(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
23 October 2003 (23.10.2003)

PCT

(10) International Publication Number
WO 03/087338 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US03/11392

(22) International Filing Date: 11 April 2003 (11.04.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/371,912 11 April 2002 (11.04.2002) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/371,912 (CON)
Filed on 11 April 2002 (11.04.2002)

(71) Applicant *(for all designated States except US)*: **AMGEN, INC.** [US/US]; One Amgen Center Drive, Thousand Oaks, CA 91320-1799 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **TATAREWICZ, Suzanna** [US/US]; 281 Brook Road, Newbury Park, CA 91320 (US). **JING, Shuqian** [US/US]; 3254 Bordero Lane, Thousand Oaks, CA 91362 (US).

(74) Agent: **ZUHN, Donald, L.**; McDonnell Boehnen Hulbert & Berghoff, 300 South Wacker Drive, Chicago, IL 60606 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/087338 A2

(54) Title: HER-2 RECEPTOR TYROSINE KINASE MOLECULES AND USES THEREOF

(57) Abstract: The present invention provides HER-2 Receptor Tyrosine Kinase polypeptides and nucleic acid molecules encoding the same. Specifically, the present invention relates to splice variants of HER-2 (HER-2sv). The invention also provides selective binding agents, vectors, host cells, and methods for producing HER-2sv polypeptides. The invention further provides pharmaceutical compositions and methods for the diagnosis, treatment, amelioration, and/or prevention of diseases, disorders, and conditions associated with HER-2sv polypeptides.

HER-2 RECEPTOR TYROSINE KINASE MOLECULES AND USES THEREOF

This application claims the benefit of priority from U.S. Provisional App. No. 60/371,912, filed April 11, 2002, the disclosure of which is explicitly incorporated by reference herein.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to HER-2 Receptor Tyrosine Kinase polypeptides and nucleic acid molecules encoding the same. Specifically, the present invention relates to splice variants of HER-2 (HER-2sv). The invention also relates to selective binding agents, vectors, host cells, and methods for producing HER-2sv polypeptides. The invention further relates to pharmaceutical compositions and methods for the diagnosis, treatment, amelioration, and prevention of diseases, disorders, and conditions associated with HER-2sv polypeptides.

2. Background of the Invention

Technical advances in the identification, cloning, expression, and manipulation of nucleic acid molecules and the deciphering of the human genome have greatly accelerated the discovery of novel therapeutics. Rapid nucleic acid sequencing techniques can now generate sequence information at unprecedented rates and, coupled with computational analyses, allow the assembly of overlapping sequences into partial and entire genomes and the identification of polypeptide-encoding regions. A comparison of a predicted amino acid sequence against a database compilation of known amino acid sequences allows one to determine the extent of homology to previously identified sequences and/or structural landmarks. The cloning and expression of a polypeptide-encoding region of a nucleic acid molecule provides a polypeptide product for structural and functional analyses. The manipulation of nucleic acid molecules and encoded polypeptides may confer advantageous properties on a product for use as a therapeutic.

In spite of the significant technical advances in genome research over the past decade, the potential for the development of novel therapeutics based on the human genome is still largely unrealized. Many genes encoding potentially beneficial polypeptide therapeutics or those encoding polypeptides, which may act as "targets" for therapeutic molecules, have still not been identified. Accordingly, it is an

object of the invention to identify novel polypeptides, and nucleic acid molecules encoding the same, which have diagnostic or therapeutic benefit.

The HER-2 (also known as erbB-2, c-neu, or HER-2/neu) proto-oncogene is a member of the epidermal growth factor (EGF) family. Other members of the EGF family include the epidermal growth factor receptor (EGFR or HER-1), ErbB-3/HER-3, and ErbB-4/HER-4. HER-2 encodes a transmembrane receptor (p185) having tyrosine kinase activity, and which is associated with multiple signal transduction pathways. Abberant HER-2 expression has been detected in many different types of human cancers, including breast, ovarian, gastric, lung, and oral cancer. HER-2 is an important prognostic and predictive factor in breast cancer in that HER-2 overexpression in breast cancer has been associated with poor overall survival and has been shown to enhance malignancy. Because this malignant phenotype can be suppressed through HER-2 repression, HER-2 is a significant target for developing antiucancer agents.

SUMMARY OF THE INVENTION

The present invention relates to novel HER-2sv nucleic acid molecules and encoded polypeptides.

The invention provides for an isolated nucleic acid molecule comprising:

(a) the nucleotide sequence as set forth in any of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9;

(b) a nucleotide sequence encoding the polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10;

(c) a nucleotide sequence that hybridizes under at least moderately stringent conditions to the complement of the nucleotide sequence of any of (a) or (b), wherein the encoded polypeptide has an activity of the polypeptide set forth in in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10; or

(d) a nucleotide sequence complementary to the nucleotide sequence of any of (a) - (c).

The invention also provides for an isolated nucleic acid molecule comprising:

(a) a nucleotide sequence encoding a polypeptide that is at least about 70 percent identical to the polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, wherein the encoded polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10;

(b) a region of the nucleotide sequence of any of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9 encoding a polypeptide fragment of at least about 25 amino acid residues, wherein the polypeptide fragment has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, or is antigenic;

(c) a region of the nucleotide sequence of SEQ ID NO: 1 encoding a polypeptide fragment of at least about 25 amino acid residues, including residues 261 through 262 of SEQ ID NO: 2, wherein the polypeptide fragment has an activity of the polypeptide set forth in SEQ ID NO: 2, or is antigenic;

(d) a region of the nucleotide sequence of SEQ ID NO: 3 encoding a polypeptide fragment of at least about 25 amino acid residues, including residues 383 through 384 of SEQ ID NO: 4, wherein the polypeptide fragment has an activity of the polypeptide set forth in SEQ ID NO: 4, or is antigenic;

(e) a region of the nucleotide sequence of SEQ ID NO: 5 encoding a polypeptide fragment of at least about 25 amino acid residues, including residues 384 through 422 of SEQ ID NO: 6, wherein the polypeptide fragment has an activity of the polypeptide set forth in SEQ ID NO: 6, or is antigenic;

(f) a region of the nucleotide sequence of SEQ ID NO: 9 encoding a polypeptide fragment of at least about 25 amino acid residues, including residues 580 through 613 of SEQ ID NO: 10, wherein the polypeptide fragment has an activity of the polypeptide set forth in SEQ ID NO: 10, or is antigenic;

(g) a nucleotide sequence that hybridizes under at least moderately stringent conditions to the complement of the nucleotide sequence of any of (a) - (f), wherein the encoded polypeptide has an activity of the polypeptide set forth in in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10; or

(h) a nucleotide sequence complementary to the nucleotide sequence of any of (a) - (g).

The invention further provides for an isolated nucleic acid molecule comprising:

(a) a nucleotide sequence encoding a polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10 with at least one conservative amino acid substitution, wherein the encoded polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10;

(b) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO: 2 having a C- and/or N- terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2, and wherein the polypeptide includes residues 261 through 262 of SEQ ID NO: 2;

(c) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO: 4 having a C- and/or N- terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 4, and wherein the polypeptide includes residues 383 through 384 of SEQ ID NO: 4;

(d) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO: 6 having a C- and/or N- terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 6, and wherein the polypeptide includes residues 384 through 422 of SEQ ID NO: 6;

(e) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO: 10 having a C- and/or N- terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 10, and wherein the polypeptide includes residues 580 through 613 of SEQ ID NO: 10;

(e) a nucleotide sequence encoding a polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10 with at least one modification that is an amino acid substitution, C-terminal truncation, or N-terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, and wherein the polypeptide includes residues 261 through 262 of SEQ ID NO: 2, residues 383 through 384 of SEQ ID NO: 4, residues 384 through 422 of SEQ ID NO: 6, or residues 580 through 613 of SEQ ID NO: 10;

(f) a nucleotide sequence that hybridizes under at least moderately stringent conditions to the complement of the nucleotide sequence of any of (a) - (e),

wherein the encoded polypeptide has an activity of the polypeptide set forth in in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10; or

(g) a nucleotide sequence complementary to the nucleotide sequence of any of (a) - (f).

The present invention provides for an isolated polypeptide comprising the amino acid as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10.

The invention also provides for an isolated polypeptide comprising:

(a) an amino acid sequence for an ortholog of any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10;

(b) an amino acid sequence that is at least about 70 percent identical to the amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10;

(c) a fragment of the amino acid sequence set forth in SEQ ID NO: 2 comprising at least about 25 amino acid residues, including residues 261 through 262 of SEQ ID NO: 2, wherein the polypeptide fragment has an activity of the polypeptide set forth in SEQ ID NO: 2, or is antigenic;

(d) a fragment of the amino acid sequence set forth in SEQ ID NO: 4 comprising at least about 25 amino acid residues, including residues 383 through 384 of SEQ ID NO: 4, wherein the polypeptide fragment has an activity of the polypeptide set forth in SEQ ID NO: 4, or is antigenic;

(e) a fragment of the amino acid sequence set forth in SEQ ID NO: 6 comprising at least about 25 amino acid residues, including residues 384 through 422 of SEQ ID NO: 6, wherein the polypeptide fragment has an activity of the polypeptide set forth in SEQ ID NO: 6, or is antigenic; or

(f) a fragment of the amino acid sequence set forth in SEQ ID NO: 10 comprising at least about 25 amino acid residues, including residues 580 through 613 of SEQ ID NO: 10, wherein the polypeptide fragment has an activity of the polypeptide set forth in SEQ ID NO: 10, or is antigenic.

The invention further provides for an isolated polypeptide comprising:

(a) the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10 with at least one conservative amino acid substitution, wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10;

(b) the amino acid sequence as set forth in SEQ ID NO: 2 having a C- and/or N- terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2, and wherein the polypeptide includes residues 261 through 262 of SEQ ID NO: 2;

(c) the amino acid sequence as set forth in SEQ ID NO: 4 having a C- and/or N- terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 4, and wherein the polypeptide includes residues 383 through 384 of SEQ ID NO: 4;

(d) the amino acid sequence as set forth in SEQ ID NO: 6 having a C- and/or N- terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 6, and wherein the polypeptide includes residues 384 through 422 of SEQ ID NO: 6;

(e) the amino acid sequence as set forth in SEQ ID NO: 10 having a C- and/or N- terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 10, and wherein the polypeptide includes residues 580 through 613 of SEQ ID NO: 10; or

(f) the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10 with at least one modification that is an amino acid substitution, C-terminal truncation, or N-terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, and wherein the polypeptide includes residues 261 through 262 of SEQ ID NO: 2, residues 383 through 384 of SEQ ID NO: 4, residues 384 through 422 of SEQ ID NO: 6, or residues 580 through 613 of SEQ ID NO: 10.

Also provided are fusion polypeptides comprising HER-2sv amino acid sequences.

The present invention also provides for an expression vector comprising the isolated nucleic acid molecules as set forth herein, recombinant host cells comprising

the recombinant nucleic acid molecules as set forth herein, and a method of producing a HER-2sv polypeptide comprising culturing the host cells and optionally isolating the polypeptide so produced. Isolation of the expressed polypeptide is described as optional because there may be instances where it is desired to express the polypeptide on the cell surface or on a cell membrane for use in screening methods for the identification of antagonists of HER-2sv activity.

A transgenic non-human animal comprising a nucleic acid molecule encoding a HER-2sv polypeptide is also encompassed by the invention. The HER-2sv nucleic acid molecules are introduced into the animal in a manner that allows expression and increased levels of a HER-2sv polypeptide, which may include increased circulating levels. Alternatively, the HER-2sv nucleic acid molecules are introduced into the animal in a manner that prevents expression of endogenous HER-2sv polypeptide (*i.e.*, generates a transgenic animal possessing a HER-2sv polypeptide gene knockout). The transgenic non-human animal is preferably a mammal, and more preferably a rodent, such as a rat or a mouse.

Also provided are derivatives of the HER-2sv polypeptides of the present invention.

Additionally provided are selective binding agents such as antibodies and peptides capable of specifically binding the HER-2sv polypeptides of the invention.

Pharmaceutical compositions comprising the nucleotides, polypeptides, or selective binding agents of the invention and one or more pharmaceutically acceptable formulation agents are also encompassed by the invention. The pharmaceutical compositions are used to provide therapeutically effective amounts of the nucleotides or polypeptides of the present invention. The invention is also directed to methods of using the polypeptides, nucleic acid molecules, and selective binding agents.

The HER-2sv polypeptides and nucleic acid molecules of the present invention may be used to treat, prevent, ameliorate, and/or detect diseases and disorders, including those recited herein.

The present invention also provides a method of assaying test molecules to identify a test molecule that binds to a HER-2sv polypeptide. The method comprises contacting a HER-2sv polypeptide with a test molecule to determine the extent of binding of the test molecule to the polypeptide. The method further comprises determining whether such test molecules are antagonists of a HER-2sv polypeptide.

The present invention further provides a method of testing the impact of molecules on the expression of HER-2sv polypeptide or on the activity of HER-2sv polypeptide.

Methods of regulating expression and modulating (*i.e.*, increasing or decreasing) levels of a HER-2sv polypeptide are also encompassed by the invention. One method comprises administering to an animal a nucleic acid molecule encoding a HER-2sv polypeptide. In another method, a nucleic acid molecule comprising elements that regulate or modulate the expression of a HER-2sv polypeptide may be administered. Examples of these methods include gene therapy, cell therapy, and anti-sense therapy as further described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-1C illustrate the nucleotide sequence of the human HER2-sv form 68 gene (SEQ ID NO: 1) and the deduced amino acid sequence of human HER2-sv form 68 polypeptide (SEQ ID NO: 2);

Figures 2A-2D illustrate the nucleotide sequence of the human HER2-sv form 97 gene (SEQ ID NO: 3) and the deduced amino acid sequence of human HER2-sv form 97 polypeptide (SEQ ID NO: 4);

Figures 3A-3D illustrate the nucleotide sequence of the human HER2-sv form 119 gene (SEQ ID NO: 5) and the deduced amino acid sequence of human HER2-sv form 119 polypeptide (SEQ ID NO: 6);

Figure 4 illustrates the nucleotide sequence of the human HER2-sv form 156 gene (SEQ ID NO: 7) and the deduced amino acid sequence of human HER2-sv form 156 polypeptide (SEQ ID NO: 8);

Figures 5A-5D illustrate the nucleotide sequence of the human HER2-sv form 184 gene (SEQ ID NO: 9) and the deduced amino acid sequence of human HER2-sv form 184 polypeptide (SEQ ID NO: 10);

Figures 6A-6C illustrate the amino acid sequence alignment of the extracellular portion of human HER-2 (SEQ ID NO: 11) and human HER-2sv forms 97 (SEQ ID

NO: 4), 184 (SEQ ID NO: 10), 119 (SEQ ID NO: 6), 68 (SEQ ID NO: 2), and 156 (SEQ ID NO: 8);

Figure 7 illustrates a schematic representation of the structure of the known form of the extracellular domain of the HER-2 gene and human HER-2sv forms 119, 184, 97, 68, and 156. The functional domains (two L-domains and a furin-like domain) in the extracellular domain of the HER-2 gene are indicated.

DETAILED DESCRIPTION OF THE INVENTION

The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described. All references cited in this application are expressly incorporated by reference herein.

1. Definitions

The terms "HER-2sv gene" or "HER-2sv nucleic acid molecule" or "HER-2sv polynucleotide" refer to a nucleic acid molecule comprising or consisting of a nucleotide sequence as set forth in any of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9; a nucleotide sequence encoding the polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10; and nucleic acid molecules as defined herein.

The term "HER-2sv polypeptide allelic variant" refers to one of several possible naturally occurring alternate forms of a gene occupying a given locus on a chromosome of an organism or a population of organisms.

The term "isolated nucleic acid molecule" refers to a nucleic acid molecule of the invention that (1) has been separated from at least about 50 percent of proteins, lipids, carbohydrates, or other materials with which it is naturally found when total nucleic acid is isolated from the source cells, (2) is not linked to all or a portion of a polynucleotide to which the "isolated nucleic acid molecule" is linked in nature, (3) is operably linked to a polynucleotide which it is not linked to in nature, or (4) does not occur in nature as part of a larger polynucleotide sequence. Preferably, the isolated nucleic acid molecule of the present invention is substantially free from any other contaminating nucleic acid molecule(s) or other contaminants that are found in its natural environment that would interfere with its use in polypeptide production or its therapeutic, diagnostic, prophylactic or research use.

The term "nucleic acid sequence" or "nucleic acid molecule" refers to a DNA or RNA sequence. The term encompasses molecules formed from any of the known base analogs of DNA and RNA such as, but not limited to 4-acetylcytosine, 8-hydroxy-N6-methyladenosine, aziridinyl-cytosine, pseudoisocytosine, 5-(carboxyhydroxylmethyl) uracil, 5-fluorouracil, 5-bromouracil, 5-carboxymethylaminomethyl-2-thiouracil, 5-carboxy-methylaminomethyluracil, dihydrouracil, inosine, N6-iso-pentenyladenine, 1-methyladenine, 1-methylpseudouracil, 1-methylguanine, 1-methylinosine, 2,2-dimethyl-guanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-methyladenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyamino-methyl-2-thiouracil, beta-D-mannosylqueosine, 5' -methoxycarbonyl-methyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, oxybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, N-uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, pseudouracil, queosine, 2-thiocytosine, and 2,6-diaminopurine.

The term "vector" is used to refer to any molecule (*e.g.*, nucleic acid, plasmid, or virus) used to transfer coding information to a host cell.

The term "expression vector" refers to a vector that is suitable for transformation of a host cell and contains nucleic acid sequences that direct and/or control the expression of inserted heterologous nucleic acid sequences. Expression includes, but is not limited to, processes such as transcription, translation, and RNA splicing, if introns are present.

The term "operably linked" is used herein to refer to an arrangement of flanking sequences wherein the flanking sequences so described are configured or assembled so as to perform their usual function. Thus, a flanking sequence operably linked to a coding sequence may be capable of effecting the replication, transcription and/or translation of the coding sequence. For example, a coding sequence is operably linked to a promoter when the promoter is capable of directing transcription of that coding sequence. A flanking sequence need not be contiguous with the coding sequence, so long as it functions correctly. Thus, for example, intervening untranslated yet transcribed sequences can be present between a promoter sequence and the coding sequence and the promoter sequence can still be considered "operably linked" to the coding sequence.

The term "host cell" is used to refer to a cell which has been transformed, or is capable of being transformed with a nucleic acid sequence and then of expressing a selected gene of interest. The term includes the progeny of the parent cell, whether or not the progeny is identical in morphology or in genetic make-up to the original parent, so long as the selected gene is present.

The term "HER-2sv polypeptide" refers to a polypeptide comprising the amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, OR SEQ ID NO: 10 and related polypeptides. Related polypeptides include HER-2sv polypeptide fragments, HER-2sv polypeptide orthologs, HER-2sv polypeptide variants, and HER-2sv polypeptide derivatives, which possess at least one activity of the polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, OR SEQ ID NO: 10. HER-2sv polypeptides may be mature polypeptides, as defined herein, and may or may not have an amino-terminal methionine residue, depending on the method by which they are prepared.

The term "HER-2sv polypeptide fragment" refers to a polypeptide that comprises a truncation at the amino-terminus (with or without a leader sequence) and/or a truncation at the carboxyl-terminus of the polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. The term "HER-2sv polypeptide fragment" also refers to amino-terminal and/or carboxyl-terminal truncations of HER-2sv polypeptide orthologs, HER-2sv polypeptide derivatives, or HER-2sv polypeptide variants, or to amino-terminal and/or carboxyl-terminal truncations of the polypeptides encoded by HER-2sv polypeptide allelic variants. HER-2sv polypeptide fragments may result from *in vivo* protease activity. Membrane-bound forms of a HER-2sv polypeptide are also contemplated by the present invention. In preferred embodiments, truncations comprise about 10 amino acids, or about 20 amino acids, or about 50 amino acids, or about 75 amino acids, or about 100 amino acids, or more than about 100 amino acids. The polypeptide fragments so produced will comprise about 25 contiguous amino acids, or about 50 amino acids, or about 75 amino acids, or about 100 amino acids, or about 150 amino acids, or about 200 amino acids, or more than about 200 amino acids. Such HER-2sv polypeptide fragments may optionally comprise an amino-terminal methionine residue. It will be appreciated that such fragments can be used, for example, to generate antibodies to HER-2sv polypeptides.

The term "HER-2sv polypeptide ortholog" refers to a polypeptide from another species that corresponds to HER-2sv polypeptide amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. For example, mouse and human HER-2sv polypeptides are considered orthologs of each other.

The term "HER-2sv polypeptide variants" refers to HER-2sv polypeptides comprising amino acid sequences having one or more amino acid sequence substitutions, deletions (such as internal deletions and/or HER-2sv polypeptide fragments), and/or additions (such as internal additions and/or HER-2sv fusion polypeptides) as compared to the HER-2sv polypeptide amino acid sequence set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10 (with or without a leader sequence). Variants may be naturally occurring (*e.g.*, HER-2sv polypeptide allelic variants and HER-2sv polypeptide orthologs) or artificially constructed. Such HER-2sv polypeptide variants may be prepared from the corresponding nucleic acid molecules having a DNA sequence that varies accordingly from the DNA sequence as set forth in any of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9. In preferred embodiments, the variants have from 1 to 3, or from 1 to 5, or from 1 to 10, or from 1 to 15, or from 1 to 20, or from 1 to 25, or from 1 to 50, or from 1 to 75, or from 1 to 100, or more than 100 amino acid substitutions, wherein the substitutions may be conservative, or non-conservative, or any combination thereof.

The term "HER-2sv polypeptide derivatives" refers to the polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, HER-2sv polypeptide fragments, HER-2sv polypeptide orthologs, or HER-2sv polypeptide variants, as defined herein, that have been chemically modified. The term "HER-2sv polypeptide derivatives" also refers to the polypeptides encoded by HER-2sv polypeptide allelic variants, as defined herein, that have been chemically modified.

The term "mature HER-2sv polypeptide" refers to a HER-2sv polypeptide lacking a leader sequence. A mature HER-2sv polypeptide may also include other modifications such as proteolytic processing of the amino-terminus (with or without a leader sequence) and/or the carboxyl-terminus, cleavage of a smaller polypeptide from a larger precursor, N-linked and/or O-linked glycosylation, and the like.

The term "HER-2sv fusion polypeptide" refers to a fusion of one or more amino acids (such as a heterologous protein or peptide) at the amino- or carboxyl-terminus of the polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, HER-2sv polypeptide fragments, HER-2sv polypeptide orthologs, HER-2sv polypeptide variants, or HER-2sv derivatives, as defined herein. The term "HER-2sv fusion polypeptide" also refers to a fusion of one or more amino acids at the amino- or carboxyl-terminus of the polypeptide encoded by HER-2sv polypeptide allelic variants, as defined herein.

The term "biologically active HER-2sv polypeptides" refers to HER-2sv polypeptides having at least one activity characteristic of the polypeptide comprising the amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. In addition, a HER-2sv polypeptide may be active as an immunogen; that is, the HER-2sv polypeptide contains at least one epitope to which antibodies may be raised.

The term "isolated polypeptide" refers to a polypeptide of the present invention that (1) has been separated from at least about 50 percent of polynucleotides, lipids, carbohydrates, or other materials with which it is naturally found when isolated from the source cell, (2) is not linked (by covalent or noncovalent interaction) to all or a portion of a polypeptide to which the "isolated polypeptide" is linked in nature, (3) is operably linked (by covalent or noncovalent interaction) to a polypeptide with which it is not linked in nature, or (4) does not occur in nature. Preferably, the isolated polypeptide is substantially free from any other contaminating polypeptides or other contaminants that are found in its natural environment that would interfere with its therapeutic, diagnostic, prophylactic or research use.

The term "identity," as known in the art, refers to a relationship between the sequences of two or more polypeptide molecules or two or more nucleic acid molecules, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between nucleic acid molecules or polypeptides, as the case may be, as determined by the match between strings of two or more nucleotide or two or more amino acid sequences. "Identity" measures the percent of identical matches between the smaller of two or more sequences with gap alignments (if any) addressed by a particular mathematical model or computer program (*i.e.*, "algorithms").

The term "similarity" is a related concept, but in contrast to "identity," "similarity" refers to a measure of relatedness that includes both identical matches and conservative substitution matches. If two polypeptide sequences have, for example, 10/20 identical amino acids, and the remainder are all non-conservative substitutions, then the percent identity and similarity would both be 50%. If in the same example, there are five more positions where there are conservative substitutions, then the percent identity remains 50%, but the percent similarity would be 75% (15/20). Therefore, in cases where there are conservative substitutions, the percent similarity between two polypeptides will be higher than the percent identity between those two polypeptides.

The term "naturally occurring" or "native" when used in connection with biological materials such as nucleic acid molecules, polypeptides, host cells, and the like, refers to materials which are found in nature and are not manipulated by man. Similarly, "non-naturally occurring" or "non-native" as used herein refers to a material that is not found in nature or that has been structurally modified or synthesized by man.

The terms "effective amount" and "therapeutically effective amount" each refer to the amount of a HER-2sv polypeptide or HER-2sv nucleic acid molecule used to support an observable level of one or more biological activities of the HER-2sv polypeptides as set forth herein.

The term "pharmaceutically acceptable carrier" or "physiologically acceptable carrier" as used herein refers to one or more formulation materials suitable for accomplishing or enhancing the delivery of the HER-2sv polypeptide, HER-2sv nucleic acid molecule, or HER-2sv selective binding agent as a pharmaceutical composition.

The term "antigen" refers to a molecule or a portion of a molecule capable of being bound by a selective binding agent, such as an antibody, and additionally capable of being used in an animal to produce antibodies capable of binding to an epitope of that antigen. An antigen may have one or more epitopes.

The term "selective binding agent" refers to a molecule or molecules having specificity for a HER-2sv polypeptide. As used herein, the terms, "specific" and "specificity" refer to the ability of the selective binding agents to bind to human HER-2sv polypeptides and not to bind to human non-HER-2sv polypeptides. It will be appreciated, however, that the selective binding agents may also bind orthologs of the

polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, that is, interspecies versions thereof, such as mouse and rat HER-2sv polypeptides.

The term "transduction" is used to refer to the transfer of genes from one bacterium to another, usually by a phage. "Transduction" also refers to the acquisition and transfer of eukaryotic cellular sequences by retroviruses.

The term "transfection" is used to refer to the uptake of foreign or exogenous DNA by a cell, and a cell has been "transfected" when the exogenous DNA has been introduced inside the cell membrane. A number of transfection techniques are well known in the art and are disclosed herein. *See, e.g.*, Graham *et al.*, 1973, *Virology* 52:456; Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (Cold Spring Harbor Laboratories, 1989); Davis *et al.*, *Basic Methods in Molecular Biology* (Elsevier, 1986); and Chu *et al.*, 1981, *Gene* 13:197. Such techniques can be used to introduce one or more exogenous DNA moieties into suitable host cells.

The term "transformation" as used herein refers to a change in a cell's genetic characteristics, and a cell has been transformed when it has been modified to contain a new DNA. For example, a cell is transformed where it is genetically modified from its native state. Following transfection or transduction, the transforming DNA may recombine with that of the cell by physically integrating into a chromosome of the cell, may be maintained transiently as an episomal element without being replicated, or may replicate independently as a plasmid. A cell is considered to have been stably transformed when the DNA is replicated with the division of the cell.

2. Relatedness of Nucleic Acid Molecules and/or Polypeptide

It is understood that related nucleic acid molecules include allelic variants of the nucleic acid molecule of any of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9, and include sequences which are complementary to any of the above nucleotide sequences. Related nucleic acid molecules also include a nucleotide sequence encoding a polypeptide comprising or consisting essentially of a substitution, modification, addition and/or deletion of one or more amino acid residues compared to the polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. Such related HER-2sv polypeptides may comprise, for example, an addition and/or a deletion of one or more

N-linked or O-linked glycosylation sites or an addition and/or a deletion of one or more cysteine residues.

Related nucleic acid molecules also include fragments of HER-2sv nucleic acid molecules which encode a polypeptide of at least about 25 contiguous amino acids, or about 50 amino acids, or about 75 amino acids, or about 100 amino acids, or about 150 amino acids, or about 200 amino acids, or more than 200 amino acid residues of the HER-2sv polypeptide of any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10.

In addition, related HER-2sv nucleic acid molecules also include those molecules which comprise nucleotide sequences which hybridize under moderately or highly stringent conditions as defined herein with the fully complementary sequence of the HER-2sv nucleic acid molecule of any of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9, or of a molecule encoding a polypeptide, which polypeptide comprises the amino acid sequence as shown in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, or of a nucleic acid fragment as defined herein, or of a nucleic acid fragment encoding a polypeptide as defined herein. Hybridization probes may be prepared using the HER-2sv sequences provided herein to screen cDNA, genomic or synthetic DNA libraries for related sequences. Regions of the DNA and/or amino acid sequence of HER-2sv polypeptide that exhibit significant identity to known sequences are readily determined using sequence alignment algorithms as described herein and those regions may be used to design probes for screening.

The term "highly stringent conditions" refers to those conditions that are designed to permit hybridization of DNA strands whose sequences are highly complementary, and to exclude hybridization of significantly mismatched DNAs. Hybridization stringency is principally determined by temperature, ionic strength, and the concentration of denaturing agents such as formamide. Examples of "highly stringent conditions" for hybridization and washing are 0.015 M sodium chloride, 0.0015 M sodium citrate at 65-68°C or 0.015 M sodium chloride, 0.0015 M sodium citrate, and 50% formamide at 42°C. *See* Sambrook, Fritsch & Maniatis, *Molecular Cloning: A Laboratory Manual* (2nd ed., Cold Spring Harbor Laboratory, 1989); Anderson *et al.*, *Nucleic Acid Hybridisation: A Practical Approach* Ch. 4 (IRL Press Limited).

More stringent conditions (such as higher temperature, lower ionic strength, higher formamide, or other denaturing agent) may also be used – however, the rate of hybridization will be affected. Other agents may be included in the hybridization and washing buffers for the purpose of reducing non-specific and/or background hybridization. Examples are 0.1% bovine serum albumin, 0.1% polyvinyl-pyrrolidone, 0.1% sodium pyrophosphate, 0.1% sodium dodecylsulfate, $NaDodSO_4$, (SDS), ficoll, Denhardt's solution, sonicated salmon sperm DNA (or another non-complementary DNA), and dextran sulfate, although other suitable agents can also be used. The concentration and types of these additives can be changed without substantially affecting the stringency of the hybridization conditions. Hybridization experiments are usually carried out at pH 6.8-7.4; however, at typical ionic strength conditions, the rate of hybridization is nearly independent of pH. *See* Anderson *et al.*, *Nucleic Acid Hybridisation: A Practical Approach* Ch. 4 (IRL Press Limited).

Factors affecting the stability of DNA duplex include base composition, length, and degree of base pair mismatch. Hybridization conditions can be adjusted by one skilled in the art in order to accommodate these variables and allow DNAs of different sequence relatedness to form hybrids. The melting temperature of a perfectly matched DNA duplex can be estimated by the following equation:

$$T_m(^\circ C) = 81.5 + 16.6(\log[Na+]) + 0.41(\%G+C) - 600/N - 0.72(\%\text{formamide})$$

where N is the length of the duplex formed, [Na+] is the molar concentration of the sodium ion in the hybridization or washing solution, %G+C is the percentage of (guanine+cytosine) bases in the hybrid. For imperfectly matched hybrids, the melting temperature is reduced by approximately 1°C for each 1% mismatch.

The term "moderately stringent conditions" refers to conditions under which a DNA duplex with a greater degree of base pair mismatching than could occur under "highly stringent conditions" is able to form. Examples of typical "moderately stringent conditions" are 0.015 M sodium chloride, 0.0015 M sodium citrate at 50-65°C or 0.015 M sodium chloride, 0.0015 M sodium citrate, and 20% formamide at 37-50°C. By way of example, "moderately stringent conditions" of 50°C in 0.015 M sodium ion will allow about a 21% mismatch.

It will be appreciated by those skilled in the art that there is no absolute distinction between "highly stringent conditions" and "moderately stringent conditions." For example, at 0.015 M sodium ion (no formamide), the melting temperature of perfectly matched long DNA is about 71°C. With a wash at 65°C (at

the same ionic strength), this would allow for approximately a 6% mismatch. To capture more distantly related sequences, one skilled in the art can simply lower the temperature or raise the ionic strength.

A good estimate of the melting temperature in 1M NaCl* for oligonucleotide probes up to about 20nt is given by:

Tm = 2°C per A-T base pair + 4°C per G-C base pair

*The sodium ion concentration in 6X salt sodium citrate (SSC) is 1M. *See* Suggs *et al., Developmental Biology Using Purified Genes* 683 (Brown and Fox, eds., 1981).

High stringency washing conditions for oligonucleotides are usually at a temperature of 0-5°C below the Tm of the oligonucleotide in 6X SSC, 0.1% SDS.

In another embodiment, related nucleic acid molecules comprise or consist of a nucleotide sequence that is at least about 70 percent identical to the nucleotide sequence as shown in any of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9. In preferred embodiments, the nucleotide sequences are about 75 percent, or about 80 percent, or about 85 percent, or about 90 percent, or about 95, 96, 97, 98, or 99 percent identical to the nucleotide sequence as shown in any of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9. Related nucleic acid molecules encode polypeptides possessing at least one activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10.

Differences in the nucleic acid sequence may result in conservative and/or non-conservative modifications of the amino acid sequence relative to the amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10.

Conservative modifications to the amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10 (and the corresponding modifications to the encoding nucleotides) will produce a polypeptide having functional and chemical characteristics similar to those of HER-2sv polypeptides. In contrast, substantial modifications in the functional and/or chemical characteristics of HER-2sv polypeptides may be accomplished by selecting substitutions in the amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10 that differ significantly in their effect on maintaining (a) the structure of the molecular backbone in the area of the

substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain.

For example, a "conservative amino acid substitution" may involve a substitution of a native amino acid residue with a nonnative residue such that there is little or no effect on the polarity or charge of the amino acid residue at that position. Furthermore, any native residue in the polypeptide may also be substituted with alanine, as has been previously described for "alanine scanning mutagenesis."

Conservative amino acid substitutions also encompass non-naturally occurring amino acid residues that are typically incorporated by chemical peptide synthesis rather than by synthesis in biological systems. These include peptidomimetics, and other reversed or inverted forms of amino acid moieties.

Naturally occurring residues may be divided into classes based on common side chain properties:

1) hydrophobic: norleucine, Met, Ala, Val, Leu, Ile;

2) neutral hydrophilic: Cys, Ser, Thr;

3) acidic: Asp, Glu;

4) basic: Asn, Gln, His, Lys, Arg;

5) residues that influence chain orientation: Gly, Pro; and

6) aromatic: Trp, Tyr, Phe.

For example, non-conservative substitutions may involve the exchange of a member of one of these classes for a member from another class. Such substituted residues may be introduced into regions of the human HER-2sv polypeptide that are homologous with non-human HER-2sv polypeptides, or into the non-homologous regions of the molecule.

In making such changes, the hydropathic index of amino acids may be considered. Each amino acid has been assigned a hydropathic index on the basis of its hydrophobicity and charge characteristics. The hydropathic indices are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5).

The importance of the hydropathic amino acid index in conferring interactive biological function on a protein is generally understood in the art (Kyte *et al.*, 1982, *J.*

Mol. Biol. 157:105-31). It is known that certain amino acids may be substituted for other amino acids having a similar hydropathic index or score and still retain a similar biological activity. In making changes based upon the hydropathic index, the substitution of amino acids whose hydropathic indices are within ±2 is preferred, those that are within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred.

It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity, particularly where the biologically functionally equivalent protein or peptide thereby created is intended for use in immunological embodiments, as in the present case. The greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with its immunogenicity and antigenicity, *i.e.,* with a biological property of the protein.

The following hydrophilicity values have been assigned to these amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 ± 1); glutamate (+3.0 ± 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5 ± 1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); and tryptophan (-3.4). In making changes based upon similar hydrophilicity values, the substitution of amino acids whose hydrophilicity values are within ±2 is preferred, those that are within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred. One may also identify epitopes from primary amino acid sequences on the basis of hydrophilicity. These regions are also referred to as "epitopic core regions."

Desired amino acid substitutions (whether conservative or non-conservative) can be determined by those skilled in the art at the time such substitutions are desired. For example, amino acid substitutions can be used to identify important residues of the HER-2sv polypeptide, or to increase or decrease the affinity of the HER-2sv polypeptides described herein. Exemplary amino acid substitutions are set forth in Table I.

TABLE I

Amino Acid Substitutions

Original Residues	Exemplary Substitutions	Preferred Substitutions
Ala	Val, Leu, Ile	Val
Arg	Lys, Gln, Asn	Lys
Asn	Gln	Gln
Asp	Glu	Glu
Cys	Ser, Ala	Ser
Gln	Asn	Asn
Glu	Asp	Asp
Gly	Pro, Ala	Ala
His	Asn, Gln, Lys, Arg	Arg
Ile	Leu, Val, Met, Ala, Phe, Norleucine	Leu
Leu	Norleucine, Ile, Val, Met, Ala, Phe	Ile
Lys	Arg, 1,4 Diamino-butyric Acid, Gln, Asn	Arg
Met	Leu, Phe, Ile	Leu
Phe	Leu, Val, Ile, Ala, Tyr	Leu
Pro	Ala	Gly
Ser	Thr, Ala, Cys	Thr
Thr	Ser	Ser
Trp	Tyr, Phe	Tyr
Tyr	Trp, Phe, Thr, Ser	Phe
Val	Ile, Met, Leu, Phe, Ala, Norleucine	Leu

A skilled artisan will be able to determine suitable variants of the polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10 using well-known techniques. For identifying suitable areas of the molecule that may be changed without destroying biological activity, one skilled in the art may target areas not believed to be important for activity. For example, when similar polypeptides with similar activities from the same species or from other

species are known, one skilled in the art may compare the amino acid sequence of a HER-2sv polypeptide to such similar polypeptides. With such a comparison, one can identify residues and portions of the molecules that are conserved among similar polypeptides. It will be appreciated that changes in areas of the HER-2sv molecule that are not conserved relative to such similar polypeptides would be less likely to adversely affect the biological activity and/or structure of a HER-2sv polypeptide. One skilled in the art would also know that, even in relatively conserved regions, one may substitute chemically similar amino acids for the naturally occurring residues while retaining activity (conservative amino acid residue substitutions). Therefore, even areas that may be important for biological activity or for structure may be subject to conservative amino acid substitutions without destroying the biological activity or without adversely affecting the polypeptide structure.

Additionally, one skilled in the art can review structure-function studies identifying residues in similar polypeptides that are important for activity or structure. In view of such a comparison, one can predict the importance of amino acid residues in a HER-2sv polypeptide that correspond to amino acid residues that are important for activity or structure in similar polypeptides. One skilled in the art may opt for chemically similar amino acid substitutions for such predicted important amino acid residues of HER-2sv polypeptides.

One skilled in the art can also analyze the three-dimensional structure and amino acid sequence in relation to that structure in similar polypeptides. In view of such information, one skilled in the art may predict the alignment of amino acid residues of HER-2sv polypeptide with respect to its three dimensional structure. One skilled in the art may choose not to make radical changes to amino acid residues predicted to be on the surface of the protein, since such residues may be involved in important interactions with other molecules. Moreover, one skilled in the art may generate test variants containing a single amino acid substitution at each amino acid residue. The variants could be screened using activity assays known to those with skill in the art. Such variants could be used to gather information about suitable variants. For example, if one discovered that a change to a particular amino acid residue resulted in destroyed, undesirably reduced, or unsuitable activity, variants with such a change would be avoided. In other words, based on information gathered from such routine experiments, one skilled in the art can readily determine the amino

acids where further substitutions should be avoided either alone or in combination with other mutations.

A number of scientific publications have been devoted to the prediction of secondary structure. *See* Moult, 1996, *Curr. Opin. Biotechnol.* 7:422-27; Chou *et al.*, 1974, *Biochemistry* 13:222-45; Chou *et al.*, 1974, *Biochemistry* 113:211-22; Chou *et al.*, 1978, *Adv. Enzymol. Relat. Areas Mol. Biol.* 47:45-48; Chou *et al.*, 1978, *Ann. Rev. Biochem.* 47:251-276; and Chou *et al.*, 1979, *Biophys. J.* 26:367-84. Moreover, computer programs are currently available to assist with predicting secondary structure. One method of predicting secondary structure is based upon homology modeling. For example, two polypeptides or proteins that have a sequence identity of greater than 30%, or similarity greater than 40%, often have similar structural topologies. The recent growth of the protein structural database (PDB) has provided enhanced predictability of secondary structure, including the potential number of folds within the structure of a polypeptide or protein. *See* Holm *et al.*, 1999, *Nucleic Acids Res.* 27:244-47. It has been suggested that there are a limited number of folds in a given polypeptide or protein and that once a critical number of structures have been resolved, structural prediction will become dramatically more accurate (Brenner *et al.*, 1997, *Curr. Opin. Struct. Biol.* 7:369-76).

Additional methods of predicting secondary structure include "threading" (Jones, 1997, *Curr. Opin. Struct. Biol.* 7:377-87; Sippl *et al.*, 1996, *Structure* 4:15-19), "profile analysis" (Bowie *et al.*, 1991, *Science*, 253:164-70; Gribskov *et al.*, 1990, *Methods Enzymol.* 183:146-59; Gribskov *et al.*, 1987, *Proc. Nat. Acad. Sci. U.S.A.* 84:4355-58), and "evolutionary linkage" (*See* Holm *et al.*, *supra*, and Brenner *et al.*, *supra*).

Preferred HER-2sv polypeptide variants include glycosylation variants wherein the number and/or type of glycosylation sites have been altered compared to the amino acid sequence set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. In one embodiment, HER-2sv polypeptide variants comprise a greater or a lesser number of N-linked glycosylation sites than the amino acid sequence set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. An N-linked glycosylation site is characterized by the sequence: Asn-X-Ser or Asn-X-Thr, wherein the amino acid residue designated as X may be any amino acid residue except proline. The substitution of amino acid residues to create this sequence provides a potential new

site for the addition of an N-linked carbohydrate chain. Alternatively, substitutions that eliminate this sequence will remove an existing N-linked carbohydrate chain. Also provided is a rearrangement of N-linked carbohydrate chains wherein one or more N-linked glycosylation sites (typically those that are naturally occurring) are eliminated and one or more new N-linked sites are created. Additional preferred HER-2sv variants include cysteine variants, wherein one or more cysteine residues are deleted or substituted with another amino acid (*e.g.*, serine) as compared to the amino acid sequence set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. Cysteine variants are useful when HER-2sv polypeptides must be refolded into a biologically active conformation such as after the isolation of insoluble inclusion bodies. Cysteine variants generally have fewer cysteine residues than the native protein, and typically have an even number to minimize interactions resulting from unpaired cysteines.

In other embodiments, HER-2sv polypeptide variants comprise an amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10 with at least one amino acid insertion and wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, or an amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10 with at least one amino acid deletion and wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. HER-2sv polypeptide variants also comprise an amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10 wherein the polypeptide has a carboxyl- and/or amino-terminal truncation and further wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. HER-2sv polypeptide variants further comprise an amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10 with at least one modification that is an amino acid substitution, an amino acid insertion, an amino acid deletion, carboxyl-terminal truncation, or amino-terminal truncation and wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10.

In further embodiments, HER-2sv polypeptide variants comprise an amino acid sequence that is at least about 70 percent identical to the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. In preferred embodiments, HER-2sv polypeptide variants comprise an amino acid sequence that is at least about 75 percent, or about 80 percent, or about 85 percent, or about 90 percent, or about 95, 96, 97, 98, or 99 percent identical percent to the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. HER-2sv polypeptide variants possess at least one activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10.

In addition, the polypeptide comprising the amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, or other HER-2sv polypeptide, may be fused to a homologous polypeptide to form a homodimer or to a heterologous polypeptide to form a heterodimer. Heterologous peptides and polypeptides include, but are not limited to: an epitope to allow for the detection and/or isolation of a HER-2sv fusion polypeptide; a transmembrane receptor protein or a portion thereof, such as an extracellular domain or a transmembrane and intracellular domain; a ligand or a portion thereof which binds to a transmembrane receptor protein; an enzyme or portion thereof which is catalytically active; a polypeptide or peptide which promotes oligomerization, such as a leucine zipper domain; a polypeptide or peptide which increases stability, such as an immunoglobulin constant region; and a polypeptide which has a therapeutic activity different from the polypeptide comprising the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, or other HER-2sv polypeptide.

Fusions can be made either at the amino-terminus or at the carboxyl-terminus of the polypeptide comprising the amino acid sequence set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, or other HER-2sv polypeptide. Fusions may be direct with no linker or adapter molecule or may be through a linker or adapter molecule. A linker or adapter molecule may be one or more amino acid residues, typically from about 20 to about 50 amino acid residues. A linker or adapter molecule may also be designed with a cleavage site for a DNA restriction endonuclease or for a protease to allow for the separation of the

fused moieties. It will be appreciated that once constructed, the fusion polypeptides can be derivatized according to the methods described herein.

In a further embodiment of the invention, the polypeptide comprising the amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, or other HER-2sv polypeptide, is fused to one or more domains of an Fc region of human IgG. Antibodies comprise two functionally independent parts, a variable domain known as "Fab," that binds an antigen, and a constant domain known as "Fc," that is involved in effector functions such as complement activation and attack by phagocytic cells. An Fc has a long serum half-life, whereas an Fab is short-lived. Capon *et al.*, 1989, *Nature* 337:525-31. When constructed together with a therapeutic protein, an Fc domain can provide longer half-life or incorporate such functions as Fc receptor binding, protein A binding, complement fixation, and perhaps even placental transfer. *Id.* Table II summarizes the use of certain Fc fusions known in the art.

TABLE II

Fc Fusion with Therapeutic Proteins

Form of Fc	Fusion partner	Therapeutic implications	Reference
IgG1	N-terminus of CD30-L	Hodgkin's disease; anaplastic lymphoma; T-cell leukemia	U.S. Patent No. 5,480,981
Murine Fcγ2a	IL-10	anti-inflammatory; transplant rejection	Zheng *et al.*, 1995, *J. Immunol.* 154:5590-600
IgG1	TNF receptor	septic shock	Fisher *et al.*, 1996, *N. Engl. J. Med.* 334:1697-1702; Van Zee *et al.*, 1996, *J. Immunol.* 156:2221-30
IgG, IgA, IgM, or IgE (excluding the first domain)	TNF receptor	inflammation, autoimmune disorders	U.S. Patent No. 5,808,029
IgG1	CD4 receptor	AIDS	Capon *et al.*, 1989, *Nature* 337: 525-31
IgG1, IgG3	N-terminus of IL-2	anti-cancer, antiviral	Harvill *et al.*, 1995, *Immunotech.* 1:95-105
IgG1	C-terminus of OPG	osteoarthritis; bone density	WO 97/23614
IgG1	N-terminus of	anti-obesity	PCT/US 97/23183, filed

	leptin		December 11, 1997
Human Ig Cγ1	CTLA-4	autoimmune disorders	Linsley, 1991, *J. Exp. Med.*, 174:561-69

In one example, a human IgG hinge, CH2, and CH3 region may be fused at either the amino-terminus or carboxyl-terminus of the HER-2sv polypeptides using methods known to the skilled artisan. In another example, a human IgG hinge, CH2, and CH3 region may be fused at either the amino-terminus or carboxyl-terminus of a HER-2sv polypeptide fragment (*e.g.*, the predicted extracellular portion of HER-2sv polypeptide).

The resulting HER-2sv fusion polypeptide may be purified by use of a Protein A affinity column. Peptides and proteins fused to an Fc region have been found to exhibit a substantially greater half-life *in vivo* than the unfused counterpart. Also, a fusion to an Fc region allows for dimerization/multimerization of the fusion polypeptide. The Fc region may be a naturally occurring Fc region, or may be altered to improve certain qualities, such as therapeutic qualities, circulation time, or reduced aggregation.

Identity and similarity of related nucleic acid molecules and polypeptides are readily calculated by known methods. Such methods include, but are not limited to those described in *Computational Molecular Biology* (A.M. Lesk, ed., Oxford University Press 1988); *Biocomputing: Informatics and Genome Projects* (D.W. Smith, ed., Academic Press 1993); *Computer Analysis of Sequence Data* (Part 1, A.M. Griffin and H.G. Griffin, eds., Humana Press 1994); G. von Heinle, *Sequence Analysis in Molecular Biology* (Academic Press 1987); *Sequence Analysis Primer* (M. Gribskov and J. Devereux, eds., M. Stockton Press 1991); and Carillo *et al.*, 1988, *SIAM J. Applied Math.*, 48:1073.

Preferred methods to determine identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are described in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package, including GAP (Devereux *et al.*, 1984, *Nucleic Acids Res.* 12:387; Genetics Computer Group, University of Wisconsin, Madison, WI), BLASTP, BLASTN, and FASTA (Altschul *et al.*, 1990, *J. Mol. Biol.* 215:403-10). The BLASTX program is publicly available from the

National Center for Biotechnology Information (NCBI) and other sources (Altschul *et al.*, *BLAST Manual* (NCB NLM NIH, Bethesda, MD); Altschul *et al.*, 1990, *supra*). The well-known Smith Waterman algorithm may also be used to determine identity.

Certain alignment schemes for aligning two amino acid sequences may result in the matching of only a short region of the two sequences, and this small aligned region may have very high sequence identity even though there is no significant relationship between the two full-length sequences. Accordingly, in a preferred embodiment, the selected alignment method (GAP program) will result in an alignment that spans at least 50 contiguous amino acids of the claimed polypeptide.

For example, using the computer algorithm GAP (Genetics Computer Group, University of Wisconsin, Madison, WI), two polypeptides for which the percent sequence identity is to be determined are aligned for optimal matching of their respective amino acids (the "matched span," as determined by the algorithm). A gap opening penalty (which is calculated as 3X the average diagonal; the "average diagonal" is the average of the diagonal of the comparison matrix being used; the "diagonal" is the score or number assigned to each perfect amino acid match by the particular comparison matrix) and a gap extension penalty (which is usually 0.1X the gap opening penalty), as well as a comparison matrix such as PAM 250 or BLOSUM 62 are used in conjunction with the algorithm. A standard comparison matrix is also used by the algorithm (*see* Dayhoff *et al.*, 5 *Atlas of Protein Sequence and Structure* (Supp. 3 1978)(PAM250 comparison matrix); Henikoff *et al.*, 1992, *Proc. Natl. Acad. Sci USA* 89:10915-19 (BLOSUM 62 comparison matrix)).

Preferred parameters for polypeptide sequence comparison include the following:

Algorithm: Needleman and Wunsch, 1970, *J. Mol. Biol.* 48:443-53;
Comparison matrix: BLOSUM 62 (Henikoff *et al.*, *supra*);
Gap Penalty: 12
Gap Length Penalty: 4
Threshold of Similarity: 0

The GAP program is useful with the above parameters. The aforementioned parameters are the default parameters for polypeptide comparisons (along with no penalty for end gaps) using the GAP algorithm.

Preferred parameters for nucleic acid molecule sequence comparison include the following:

Algorithm: Needleman and Wunsch, *supra*;
Comparison matrix: matches = +10, mismatch = 0
Gap Penalty: 50
Gap Length Penalty: 3

The GAP program is also useful with the above parameters. The aforementioned parameters are the default parameters for nucleic acid molecule comparisons.

Other exemplary algorithms, gap opening penalties, gap extension penalties, comparison matrices, and thresholds of similarity may be used, including those set forth in the Program Manual, Wisconsin Package, Version 9, September, 1997. The particular choices to be made will be apparent to those of skill in the art and will depend on the specific comparison to be made, such as DNA-to-DNA, protein-to-protein, protein-to-DNA; and additionally, whether the comparison is between given pairs of sequences (in which case GAP or BestFit are generally preferred) or between one sequence and a large database of sequences (in which case FASTA or BLASTA are preferred).

3. Nucleic Acid Molecules

The nucleic acid molecules encoding a polypeptide comprising the amino acid sequence of a HER-2sv polypeptide can readily be obtained in a variety of ways including, without limitation, chemical synthesis, cDNA or genomic library screening, expression library screening, and/or PCR amplification of cDNA.

Recombinant DNA methods used herein are generally those set forth in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 1989) and/or *Current Protocols in Molecular Biology* (Ausubel *et al.*, eds., Green Publishers Inc. and Wiley and Sons 1994). The invention provides for nucleic acid molecules as described herein and methods for obtaining such molecules.

Where a gene encoding the amino acid sequence of a HER-2sv polypeptide has been identified from one species, all or a portion of that gene may be used as a probe to identify orthologs or related genes from the same species. The probes or

primers may be used to screen cDNA libraries from various tissue sources believed to express the HER-2sv polypeptide. In addition, part or all of a nucleic acid molecule having the sequence as set forth in any of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, OR SEQ ID NO: 9 may be used to screen a genomic library to identify and isolate a gene encoding the amino acid sequence of a HER-2sv polypeptide. Typically, conditions of moderate or high stringency will be employed for screening to minimize the number of false positives obtained from the screening.

Nucleic acid molecules encoding the amino acid sequence of HER-2sv polypeptides may also be identified by expression cloning which employs the detection of positive clones based upon a property of the expressed protein. Typically, nucleic acid libraries are screened by the binding an antibody or other binding partner (*e.g.*, receptor or ligand) to cloned proteins that are expressed and displayed on a host cell surface. The antibody or binding partner is modified with a detectable label to identify those cells expressing the desired clone.

Recombinant expression techniques conducted in accordance with the descriptions set forth below may be followed to produce these polynucleotides and to express the encoded polypeptides. For example, by inserting a nucleic acid sequence that encodes the amino acid sequence of a HER-2sv polypeptide into an appropriate vector, one skilled in the art can readily produce large quantities of the desired nucleotide sequence. The sequences can then be used to generate detection probes or amplification primers. Alternatively, a polynucleotide encoding the amino acid sequence of a HER-2sv polypeptide can be inserted into an expression vector. By introducing the expression vector into an appropriate host, the encoded HER-2sv polypeptide may be produced in large amounts.

Another method for obtaining a suitable nucleic acid sequence is the polymerase chain reaction (PCR). In this method, cDNA is prepared from poly(A)+RNA or total RNA using the enzyme reverse transcriptase. Two primers, typically complementary to two separate regions of cDNA encoding the amino acid sequence of a HER-2sv polypeptide, are then added to the cDNA along with a polymerase such as *Taq* polymerase, and the polymerase amplifies the cDNA region between the two primers.

Another means of preparing a nucleic acid molecule encoding the amino acid sequence of a HER-2sv polypeptide is chemical synthesis using methods well known to the skilled artisan such as those described by Engels *et al.*, 1989, *Angew. Chem.*

Intl. Ed. 28:716-34. These methods include, *inter alia*, the phosphotriester, phosphoramidite, and H-phosphonate methods for nucleic acid synthesis. A preferred method for such chemical synthesis is polymer-supported synthesis using standard phosphoramidite chemistry. Typically, the DNA encoding the amino acid sequence of a HER-2sv polypeptide will be several hundred nucleotides in length. Nucleic acids larger than about 100 nucleotides can be synthesized as several fragments using these methods. The fragments can then be ligated together to form the full-length nucleotide sequence of a HER-2sv gene. Usually, the DNA fragment encoding the amino-terminus of the polypeptide will have an ATG, which encodes a methionine residue. This methionine may or may not be present on the mature form of the HER-2sv polypeptide, depending on whether the polypeptide produced in the host cell is designed to be secreted from that cell. Other methods known to the skilled artisan may be used as well.

In certain embodiments, nucleic acid variants contain codons which have been altered for optimal expression of a HER-2sv polypeptide in a given host cell. Particular codon alterations will depend upon the HER-2sv polypeptide and host cell selected for expression. Such "codon optimization" can be carried out by a variety of methods, for example, by selecting codons which are preferred for use in highly expressed genes in a given host cell. Computer algorithms which incorporate codon frequency tables such as "Eco_high.Cod" for codon preference of highly expressed bacterial genes may be used and are provided by the University of Wisconsin Package Version 9.0 (Genetics Computer Group, Madison, WI). Other useful codon frequency tables include "Celegans_high.cod," "Celegans_low.cod," "Drosophila_high.cod," "Human_high.cod," "Maize_high.cod," and "Yeast_high.cod."

In some cases, it may be desirable to prepare nucleic acid molecules encoding HER-2sv polypeptide variants. Nucleic acid molecules encoding variants may be produced using site directed mutagenesis, PCR amplification, or other appropriate methods, where the primer(s) have the desired point mutations (*see* Sambrook *et al.*, *supra*, and Ausubel *et al.*, *supra*, for descriptions of mutagenesis techniques). Chemical synthesis using methods described by Engels *et al.*, *supra*, may also be used to prepare such variants. Other methods known to the skilled artisan may be used as well.

4. Vectors and Host Cells

A nucleic acid molecule encoding the amino acid sequence of a HER-2sv polypeptide is inserted into an appropriate expression vector using standard ligation techniques. The vector is typically selected to be functional in the particular host cell employed (*i.e.*, the vector is compatible with the host cell machinery such that amplification of the gene and/or expression of the gene can occur). A nucleic acid molecule encoding the amino acid sequence of a HER-2sv polypeptide may be amplified/expressed in prokaryotic, yeast, insect (baculovirus systems) and/or eukaryotic host cells. Selection of the host cell will depend in part on whether a HER-2sv polypeptide is to be post-translationally modified (*e.g.*, glycosylated and/or phosphorylated). If so, yeast, insect, or mammalian host cells are preferable. For a review of expression vectors, *see Meth. Enz.*, vol. 185 (D.V. Goeddel, ed., Academic Press 1990).

Typically, expression vectors used in any of the host cells will contain sequences for plasmid maintenance and for cloning and expression of exogenous nucleotide sequences. Such sequences, collectively referred to as "flanking sequences" in certain embodiments will typically include one or more of the following nucleotide sequences: a promoter, one or more enhancer sequences, an origin of replication, a transcriptional termination sequence, a complete intron sequence containing a donor and acceptor splice site, a sequence encoding a leader sequence for polypeptide secretion, a ribosome binding site, a polyadenylation sequence, a polylinker region for inserting the nucleic acid encoding the polypeptide to be expressed, and a selectable marker element. Each of these sequences is discussed below.

Optionally, the vector may contain a "tag"-encoding sequence, *i.e.*, an oligonucleotide molecule located at the 5' or 3' end of the HER-2sv polypeptide coding sequence; the oligonucleotide sequence encodes polyHis (such as hexaHis), or another "tag" such as FLAG, HA (hemaglutinin influenza virus), or *myc* for which commercially available antibodies exist. This tag is typically fused to the polypeptide upon expression of the polypeptide, and can serve as a means for affinity purification of the HER-2sv polypeptide from the host cell. Affinity purification can be accomplished, for example, by column chromatography using antibodies against the tag as an affinity matrix. Optionally, the tag can subsequently be removed from the

purified HER-2sv polypeptide by various means such as using certain peptidases for cleavage.

Flanking sequences may be homologous (*i.e.*, from the same species and/or strain as the host cell), heterologous (*i.e.*, from a species other than the host cell species or strain), hybrid (*i.e.*, a combination of flanking sequences from more than one source), or synthetic, or the flanking sequences may be native sequences that normally function to regulate HER-2sv polypeptide expression. As such, the source of a flanking sequence may be any prokaryotic or eukaryotic organism, any vertebrate or invertebrate organism, or any plant, provided that the flanking sequence is functional in, and can be activated by, the host cell machinery.

Flanking sequences useful in the vectors of this invention may be obtained by any of several methods well known in the art. Typically, flanking sequences useful herein – other than the HER-2sv gene flanking sequences – will have been previously identified by mapping and/or by restriction endonuclease digestion and can thus be isolated from the proper tissue source using the appropriate restriction endonucleases. In some cases, the full nucleotide sequence of a flanking sequence may be known. Here, the flanking sequence may be synthesized using the methods described herein for nucleic acid synthesis or cloning.

Where all or only a portion of the flanking sequence is known, it may be obtained using PCR and/or by screening a genomic library with a suitable oligonucleotide and/or flanking sequence fragment from the same or another species. Where the flanking sequence is not known, a fragment of DNA containing a flanking sequence may be isolated from a larger piece of DNA that may contain, for example, a coding sequence or even another gene or genes. Isolation may be accomplished by restriction endonuclease digestion to produce the proper DNA fragment followed by isolation using agarose gel purification, Qiagen® column chromatography (Chatsworth, CA), or other methods known to the skilled artisan. The selection of suitable enzymes to accomplish this purpose will be readily apparent to one of ordinary skill in the art.

An origin of replication is typically a part of those prokaryotic expression vectors purchased commercially, and the origin aids in the amplification of the vector in a host cell. Amplification of the vector to a certain copy number can, in some cases, be important for the optimal expression of a HER-2sv polypeptide. If the vector of choice does not contain an origin of replication site, one may be chemically

synthesized based on a known sequence, and ligated into the vector. For example, the origin of replication from the plasmid pBR322 (New England Biolabs, Beverly, MA) is suitable for most gram-negative bacteria and various origins (*e.g.*, SV40, polyoma, adenovirus, vesicular stomatitus virus (VSV), or papillomaviruses such as HPV or BPV) are useful for cloning vectors in mammalian cells. Generally, the origin of replication component is not needed for mammalian expression vectors (for example, the SV40 origin is often used only because it contains the early promoter).

A transcription termination sequence is typically located 3' of the end of a polypeptide coding region and serves to terminate transcription. Usually, a transcription termination sequence in prokaryotic cells is a G-C rich fragment followed by a poly-T sequence. While the sequence is easily cloned from a library or even purchased commercially as part of a vector, it can also be readily synthesized using methods for nucleic acid synthesis such as those described herein.

A selectable marker gene element encodes a protein necessary for the survival and growth of a host cell grown in a selective culture medium. Typical selection marker genes encode proteins that (a) confer resistance to antibiotics or other toxins, *e.g.*, ampicillin, tetracycline, or kanamycin for prokaryotic host cells; (b) complement auxotrophic deficiencies of the cell; or (c) supply critical nutrients not available from complex media. Preferred selectable markers are the kanamycin resistance gene, the ampicillin resistance gene, and the tetracycline resistance gene. A neomycin resistance gene may also be used for selection in prokaryotic and eukaryotic host cells.

Other selection genes may be used to amplify the gene that will be expressed. Amplification is the process wherein genes that are in greater demand for the production of a protein critical for growth are reiterated in tandem within the chromosomes of successive generations of recombinant cells. Examples of suitable selectable markers for mammalian cells include dihydrofolate reductase (DHFR) and thymidine kinase. The mammalian cell transformants are placed under selection pressure wherein only the transformants are uniquely adapted to survive by virtue of the selection gene present in the vector. Selection pressure is imposed by culturing the transformed cells under conditions in which the concentration of selection agent in the medium is successively changed, thereby leading to the amplification of both the selection gene and the DNA that encodes a HER-2sv polypeptide. As a result,

increased quantities of HER-2sv polypeptide are synthesized from the amplified DNA.

A ribosome binding site is usually necessary for translation initiation of mRNA and is characterized by a Shine-Dalgarno sequence (prokaryotes) or a Kozak sequence (eukaryotes). The element is typically located 3' to the promoter and 5' to the coding sequence of a HER-2sv polypeptide to be expressed. The Shine-Dalgarno sequence is varied but is typically a polypurine (*i.e.*, having a high A-G content). Many Shine-Dalgarno sequences have been identified, each of which can be readily synthesized using methods set forth herein and used in a prokaryotic vector.

A leader, or signal, sequence may be used to direct a HER-2sv polypeptide out of the host cell. Typically, a nucleotide sequence encoding the signal sequence is positioned in the coding region of a HER-2sv nucleic acid molecule, or directly at the 5' end of a HER-2sv polypeptide coding region. Many signal sequences have been identified, and any of those that are functional in the selected host cell may be used in conjunction with a HER-2sv nucleic acid molecule. Therefore, a signal sequence may be homologous (naturally occurring) or heterologous to the HER-2sv nucleic acid molecule. Additionally, a signal sequence may be chemically synthesized using methods described herein. In most cases, the secretion of a HER-2sv polypeptide from the host cell via the presence of a signal peptide will result in the removal of the signal peptide from the secreted HER-2sv polypeptide. The signal sequence may be a component of the vector, or it may be a part of a HER-2sv nucleic acid molecule that is inserted into the vector.

Included within the scope of this invention is the use of either a nucleotide sequence encoding a native HER-2sv polypeptide signal sequence joined to a HER-2sv polypeptide coding region or a nucleotide sequence encoding a heterologous signal sequence joined to a HER-2sv polypeptide coding region. The heterologous signal sequence selected should be one that is recognized and processed, *i.e.*, cleaved by a signal peptidase, by the host cell. For prokaryotic host cells that do not recognize and process the native HER-2sv polypeptide signal sequence, the signal sequence is substituted by a prokaryotic signal sequence selected, for example, from the group of the alkaline phosphatase, penicillinase, or heat-stable enterotoxin II leaders. For yeast secretion, the native HER-2sv polypeptide signal sequence may be substituted by the yeast invertase, alpha factor, or acid phosphatase leaders. In

mammalian cell expression the native signal sequence is satisfactory, although other mammalian signal sequences may be suitable.

In some cases, such as where glycosylation is desired in a eukaryotic host cell expression system, one may manipulate the various presequences to improve glycosylation or yield. For example, one may alter the peptidase cleavage site of a particular signal peptide, or add pro-sequences, which also may affect glycosylation. The final protein product may have, in the -1 position (relative to the first amino acid of the mature protein) one or more additional amino acids incident to expression, which may not have been totally removed. For example, the final protein product may have one or two amino acid residues found in the peptidase cleavage site, attached to the amino-terminus. Alternatively, use of some enzyme cleavage sites may result in a slightly truncated form of the desired HER-2sv polypeptide, if the enzyme cuts at such area within the mature polypeptide.

In many cases, transcription of a nucleic acid molecule is increased by the presence of one or more introns in the vector; this is particularly true where a polypeptide is produced in eukaryotic host cells, especially mammalian host cells. The introns used may be naturally occurring within the HER-2sv gene especially where the gene used is a full-length genomic sequence or a fragment thereof. Where the intron is not naturally occurring within the gene (as for most cDNAs), the intron may be obtained from another source. The position of the intron with respect to flanking sequences and the HER-2sv gene is generally important, as the intron must be transcribed to be effective. Thus, when a HER-2sv cDNA molecule is being transcribed, the preferred position for the intron is 3' to the transcription start site and 5' to the poly-A transcription termination sequence. Preferably, the intron or introns will be located on one side or the other (*i.e.*, 5' or 3') of the cDNA such that it does not interrupt the coding sequence. Any intron from any source, including viral, prokaryotic and eukaryotic (plant or animal) organisms, may be used to practice this invention, provided that it is compatible with the host cell into which it is inserted. Also included herein are synthetic introns. Optionally, more than one intron may be used in the vector.

The expression and cloning vectors of the present invention will typically contain a promoter that is recognized by the host organism and operably linked to the molecule encoding the HER-2sv polypeptide. Promoters are untranscribed sequences located upstream (*i.e.*, 5') to the start codon of a structural gene (generally within

about 100 to 1000 bp) that control the transcription of the structural gene. Promoters are conventionally grouped into one of two classes: inducible promoters and constitutive promoters. Inducible promoters initiate increased levels of transcription from DNA under their control in response to some change in culture conditions, such as the presence or absence of a nutrient or a change in temperature. Constitutive promoters, on the other hand, initiate continual gene product production; that is, there is little or no control over gene expression. A large number of promoters, recognized by a variety of potential host cells, are well known. A suitable promoter is operably linked to the DNA encoding HER-2sv polypeptide by removing the promoter from the source DNA by restriction enzyme digestion and inserting the desired promoter sequence into the vector. The native HER-2sv promoter sequence may be used to direct amplification and/or expression of a HER-2sv nucleic acid molecule. A heterologous promoter is preferred, however, if it permits greater transcription and higher yields of the expressed protein as compared to the native promoter, and if it is compatible with the host cell system that has been selected for use.

Promoters suitable for use with prokaryotic hosts include the beta-lactamase and lactose promoter systems; alkaline phosphatase; a tryptophan (trp) promoter system; and hybrid promoters such as the tac promoter. Other known bacterial promoters are also suitable. Their sequences have been published, thereby enabling one skilled in the art to ligate them to the desired DNA sequence, using linkers or adapters as needed to supply any useful restriction sites.

Suitable promoters for use with yeast hosts are also well known in the art. Yeast enhancers are advantageously used with yeast promoters. Suitable promoters for use with mammalian host cells are well known and include, but are not limited to, those obtained from the genomes of viruses such as polyoma virus, fowlpox virus, adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, retroviruses, hepatitis-B virus and most preferably Simian Virus 40 (SV40). Other suitable mammalian promoters include heterologous mammalian promoters, for example, heat-shock promoters and the actin promoter.

Additional promoters which may be of interest in controlling HER-2sv gene expression include, but are not limited to: the SV40 early promoter region (Bernoist and Chambon, 1981, *Nature* 290:304-10); the CMV promoter; the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto, *et al.*, 1980, *Cell* 22:787-97); the herpes thymidine kinase promoter (Wagner *et al.*, 1981, *Proc. Natl.*

Acad. Sci. U.S.A. 78:1444-45); the regulatory sequences of the metallothionine gene (Brinster *et al.*, 1982, *Nature* 296:39-42); prokaryotic expression vectors such as the beta-lactamase promoter (Villa-Kamaroff *et al.*, 1978, *Proc. Natl. Acad. Sci. U.S.A.*, 75:3727-31); or the tac promoter (DeBoer *et al.*, 1983, *Proc. Natl. Acad. Sci. U.S.A.*, 80:21-25). Also of interest are the following animal transcriptional control regions, which exhibit tissue specificity and have been utilized in transgenic animals: the elastase I gene control region which is active in pancreatic acinar cells (Swift *et al.*, 1984, *Cell* 38:639-46; Ornitz *et al.*, 1986, *Cold Spring Harbor Symp. Quant. Biol.* 50:399-409 (1986); MacDonald, 1987, *Hepatology* 7:425-515); the insulin gene control region which is active in pancreatic beta cells (Hanahan, 1985, *Nature* 315:115-22); the immunoglobulin gene control region which is active in lymphoid cells (Grosschedl *et al.*, 1984, *Cell* 38:647-58; Adames *et al.*, 1985, *Nature* 318:533-38; Alexander *et al.*, 1987, *Mol. Cell. Biol.*, 7:1436-44); the mouse mammary tumor virus control region which is active in testicular, breast, lymphoid and mast cells (Leder *et al.*, 1986, *Cell* 45:485-95); the albumin gene control region which is active in liver (Pinkert *et al.*, 1987, *Genes and Devel.* 1:268-76); the alpha-feto-protein gene control region which is active in liver (Krumlauf *et al.*, 1985, *Mol. Cell. Biol.*, 5:1639-48; Hammer *et al.*, 1987, *Science* 235:53-58); the alpha 1-antitrypsin gene control region which is active in the liver (Kelsey *et al.*, 1987, *Genes and Devel.* 1:161-71); the beta-globin gene control region which is active in myeloid cells (Mogram *et al.*, 1985, *Nature* 315:338-40; Kollias *et al.*, 1986, *Cell* 46:89-94); the myelin basic protein gene control region which is active in oligodendrocyte cells in the brain (Readhead *et al.*, 1987, *Cell* 48:703-12); the myosin light chain-2 gene control region which is active in skeletal muscle (Sani, 1985, *Nature* 314:283-86); and the gonadotropic releasing hormone gene control region which is active in the hypothalamus (Mason *et al.*, 1986, *Science* 234:1372-78).

An enhancer sequence may be inserted into the vector to increase the transcription of a DNA encoding a HER-2sv polypeptide of the present invention by higher eukaryotes. Enhancers are cis-acting elements of DNA, usually about 10-300 bp in length, that act on the promoter to increase transcription. Enhancers are relatively orientation and position independent. They have been found 5' and 3' to the transcription unit. Several enhancer sequences available from mammalian genes are known (*e.g.*, globin, elastase, albumin, alpha-feto-protein and insulin). Typically, however, an enhancer from a virus will be used. The SV40 enhancer, the

cytomegalovirus early promoter enhancer, the polyoma enhancer, and adenovirus enhancers are exemplary enhancing elements for the activation of eukaryotic promoters. While an enhancer may be spliced into the vector at a position 5' or 3' to a HER-2sv nucleic acid molecule, it is typically located at a site 5' from the promoter.

Expression vectors of the invention may be constructed from a starting vector such as a commercially available vector. Such vectors may or may not contain all of the desired flanking sequences. Where one or more of the flanking sequences described herein are not already present in the vector, they may be individually obtained and ligated into the vector. Methods used for obtaining each of the flanking sequences are well known to one skilled in the art.

Preferred vectors for practicing this invention are those that are compatible with bacterial, insect, and mammalian host cells. Such vectors include, *inter alia*, pCRII, pCR3, and pcDNA3.1 (Invitrogen, San Diego, CA), pBSII (Stratagene, La Jolla, CA), pET15 (Novagen, Madison, WI), pGEX (Pharmacia Biotech, Piscataway, NJ), pEGFP-N2 (Clontech, Palo Alto, CA), pETL (BlueBacII, Invitrogen), pDSR-alpha (International Pub. No. WO 90/14363) and pFastBacDual (Gibco-BRL, Grand Island, NY).

Additional suitable vectors include, but are not limited to, cosmids, plasmids, or modified viruses, but it will be appreciated that the vector system must be compatible with the selected host cell. Such vectors include, but are not limited to plasmids such as Bluescript® plasmid derivatives (a high copy number ColE1-based phagemid; Stratagene Cloning Systems, La Jolla CA), PCR cloning plasmids designed for cloning Taq-amplified PCR products (*e.g.*, TOPO™ TA Cloning® Kit and PCR2.1® plasmid derivatives; Invitrogen), and mammalian, yeast or virus vectors such as a baculovirus expression system (pBacPAK plasmid derivatives; Clontech).

After the vector has been constructed and a nucleic acid molecule encoding a HER-2sv polypeptide has been inserted into the proper site of the vector, the completed vector may be inserted into a suitable host cell for amplification and/or polypeptide expression. The transformation of an expression vector for a HER-2sv polypeptide into a selected host cell may be accomplished by well known methods including methods such as transfection, infection, calcium chloride, electroporation, microinjection, lipofection, DEAE-dextran method, or other known techniques. The method selected will in part be a function of the type of host cell to be used. These

methods and other suitable methods are well known to the skilled artisan, and are set forth, for example, in Sambrook *et al.*, *supra*.

Host cells may be prokaryotic host cells (such as *E. coli*) or eukaryotic host cells (such as a yeast, insect, or vertebrate cell). The host cell, when cultured under appropriate conditions, synthesizes a HER-2sv polypeptide that can subsequently be collected from the culture medium (if the host cell secretes it into the medium) or directly from the host cell producing it (if it is not secreted). The selection of an appropriate host cell will depend upon various factors, such as desired expression levels, polypeptide modifications that are desirable or necessary for activity (such as glycosylation or phosphorylation) and ease of folding into a biologically active molecule.

A number of suitable host cells are known in the art and many are available from the American Type Culture Collection (ATCC), Manassas, VA. Examples include, but are not limited to, mammalian cells, such as Chinese hamster ovary cells (CHO), CHO DHFR(-) cells (Urlaub *et al.*, 1980, *Proc. Natl. Acad. Sci. U.S.A.* 97:4216-20), human embryonic kidney (HEK) 293 or 293T cells, or 3T3 cells. The selection of suitable mammalian host cells and methods for transformation, culture, amplification, screening, product production, and purification are known in the art. Other suitable mammalian cell lines, are the monkey COS-1 and COS-7 cell lines, and the CV-1 cell line. Further exemplary mammalian host cells include primate cell lines and rodent cell lines, including transformed cell lines. Normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, as well as primary explants, are also suitable. Candidate cells may be genotypically deficient in the selection gene, or may contain a dominantly acting selection gene. Other suitable mammalian cell lines include but are not limited to, mouse neuroblastoma N2A cells, HeLa, mouse L-929 cells, 3T3 lines derived from Swiss, Balb-c or NIH mice, BHK or HaK hamster cell lines. Each of these cell lines is known by and available to those skilled in the art of protein expression.

Similarly useful as host cells suitable for the present invention are bacterial cells. For example, the various strains of *E. coli* (*e.g.*, HB101, DH5α, DH10, and MC1061) are well-known as host cells in the field of biotechnology. Various strains of *B. subtilis*, *Pseudomonas spp.*, other *Bacillus spp.*, *Streptomyces spp.*, and the like may also be employed in this method.

Many strains of yeast cells known to those skilled in the art are also available as host cells for the expression of the polypeptides of the present invention. Preferred yeast cells include, for example, *Saccharomyces cerivisae* and *Pichia pastoris*.

Additionally, where desired, insect cell systems may be utilized in the methods of the present invention. Such systems are described, for example, in Kitts *et al.*, 1993, *Biotechniques*, 14:810-17; Lucklow, 1993, *Curr. Opin. Biotechnol.* 4:564-72; and Lucklow *et al.*, 1993, *J. Virol.*, 67:4566-79. Preferred insect cells are Sf-9 and Hi5 (Invitrogen).

One may also use transgenic animals to express glycosylated HER-2sv polypeptides. For example, one may use a transgenic milk-producing animal (a cow or goat, for example) and obtain the present glycosylated polypeptide in the animal milk. One may also use plants to produce HER-2sv polypeptides, however, in general, the glycosylation occurring in plants is different from that produced in mammalian cells, and may result in a glycosylated product which is not suitable for human therapeutic use.

5. Polypeptide Production

Host cells comprising a HER-2sv polypeptide expression vector may be cultured using standard media well known to the skilled artisan. The media will usually contain all nutrients necessary for the growth and survival of the cells. Suitable media for culturing *E. coli* cells include, for example, Luria Broth (LB) and/or Terrific Broth (TB). Suitable media for culturing eukaryotic cells include Roswell Park Memorial Institute medium 1640 (RPMI 1640), Minimal Essential Medium (MEM) and/or Dulbecco's Modified Eagle Medium (DMEM), all of which may be supplemented with serum and/or growth factors as necessary for the particular cell line being cultured. A suitable medium for insect cultures is Grace's medium supplemented with yeastolate, lactalbumin hydrolysate, and/or fetal calf serum as necessary.

Typically, an antibiotic or other compound useful for selective growth of transfected or transformed cells is added as a supplement to the media. The compound to be used will be dictated by the selectable marker element present on the plasmid with which the host cell was transformed. For example, where the selectable marker element is kanamycin resistance, the compound added to the culture medium

will be kanamycin. Other compounds for selective growth include ampicillin, tetracycline, and neomycin.

The amount of a HER-2sv polypeptide produced by a host cell can be evaluated using standard methods known in the art. Such methods include, without limitation, Western blot analysis, SDS-polyacrylamide gel electrophoresis, non-denaturing gel electrophoresis, High Performance Liquid Chromatography (HPLC) separation, immunoprecipitation, and/or activity assays such as DNA binding gel shift assays.

If a HER-2sv polypeptide has been designed to be secreted from the host cells, the majority of polypeptide may be found in the cell culture medium. If however, the HER-2sv polypeptide is not secreted from the host cells, it will be present in the cytoplasm and/or the nucleus (for eukaryotic host cells) or in the cytosol (for gram-negative bacteria host cells).

For a HER-2sv polypeptide situated in the host cell cytoplasm and/or nucleus (for eukaryotic host cells) or in the cytosol (for bacterial host cells), the intracellular material (including inclusion bodies for gram-negative bacteria) can be extracted from the host cell using any standard technique known to the skilled artisan. For example, the host cells can be lysed to release the contents of the periplasm/cytoplasm by French press, homogenization, and/or sonication followed by centrifugation.

If a HER-2sv polypeptide has formed inclusion bodies in the cytosol, the inclusion bodies can often bind to the inner and/or outer cellular membranes and thus will be found primarily in the pellet material after centrifugation. The pellet material can then be treated at pH extremes or with a chaotropic agent such as a detergent, guanidine, guanidine derivatives, urea, or urea derivatives in the presence of a reducing agent such as dithiothreitol at alkaline pH or tris carboxyethyl phosphine at acid pH to release, break apart, and solubilize the inclusion bodies. The solubilized HER-2sv polypeptide can then be analyzed using gel electrophoresis, immunoprecipitation, or the like. If it is desired to isolate the HER-2sv polypeptide, isolation may be accomplished using standard methods such as those described herein and in Marston *et al.*, 1990, *Meth. Enz.*, 182:264-75.

In some cases, a HER-2sv polypeptide may not be biologically active upon isolation. Various methods for "refolding" or converting the polypeptide to its tertiary structure and generating disulfide linkages can be used to restore biological

activity. Such methods include exposing the solubilized polypeptide to a pH usually above 7 and in the presence of a particular concentration of a chaotrope. The selection of chaotrope is very similar to the choices used for inclusion body solubilization, but usually the chaotrope is used at a lower concentration and is not necessarily the same as chaotropes used for the solubilization. In most cases the refolding/oxidation solution will also contain a reducing agent or the reducing agent plus its oxidized form in a specific ratio to generate a particular redox potential allowing for disulfide shuffling to occur in the formation of the protein's cysteine bridges. Some of the commonly used redox couples include cysteine/cystamine, glutathione (GSH)/dithiobis GSH, cupric chloride, dithiothreitol(DTT)/dithiane DTT, and 2-2-mercaptoethanol(bME)/dithio-b(ME). In many instances, a cosolvent may be used or may be needed to increase the efficiency of the refolding, and the more common reagents used for this purpose include glycerol, polyethylene glycol of various molecular weights, arginine and the like.

If inclusion bodies are not formed to a significant degree upon expression of a HER-2sv polypeptide, then the polypeptide will be found primarily in the supernatant after centrifugation of the cell homogenate. The polypeptide may be further isolated from the supernatant using methods such as those described herein.

The purification of a HER-2sv polypeptide from solution can be accomplished using a variety of techniques. If the polypeptide has been synthesized such that it contains a tag such as Hexahistidine (HER-2sv polypeptide/hexaHis) or other small peptide such as FLAG (Eastman Kodak Co., New Haven, CT) or *myc* (Invitrogen) at either its carboxyl- or amino-terminus, it may be purified in a one-step process by passing the solution through an affinity column where the column matrix has a high affinity for the tag.

For example, polyhistidine binds with great affinity and specificity to nickel. Thus, an affinity column of nickel (such as the Qiagen® nickel columns) can be used for purification of HER-2sv polypeptide/polyHis. *See, e.g., Current Protocols in Molecular Biology* § 10.11.8 (Ausubel *et al.*, eds., Green Publishers Inc. and Wiley and Sons 1993).

Additionally, HER-2SV polypeptides may be purified through the use of a monoclonal antibody that is capable of specifically recognizing and binding to a HER-2sv polypeptide.

Other suitable procedures for purification include, without limitation, affinity chromatography, immunoaffinity chromatography, ion exchange chromatography, molecular sieve chromatography, HPLC, electrophoresis (including native gel electrophoresis) followed by gel elution, and preparative isoelectric focusing ("Isoprime" machine/technique, Hoefer Scientific, San Francisco, CA). In some cases, two or more purification techniques may be combined to achieve increased purity.

HER-2sv polypeptides may also be prepared by chemical synthesis methods (such as solid phase peptide synthesis) using techniques known in the art such as those set forth by Merrifield *et al.*, 1963, *J. Am. Chem. Soc.* 85:2149; Houghten *et al.*, 1985, *Proc Natl Acad. Sci. USA* 82:5132; and Stewart and Young, *Solid Phase Peptide Synthesis* (Pierce Chemical Co. 1984). Such polypeptides may be synthesized with or without a methionine on the amino-terminus. Chemically synthesized HER-2sv polypeptides may be oxidized using methods set forth in these references to form disulfide bridges. Chemically synthesized HER-2sv polypeptides are expected to have comparable biological activity to the corresponding HER-2sv polypeptides produced recombinantly or purified from natural sources, and thus may be used interchangeably with a recombinant or natural HER-2sv polypeptide.

Another means of obtaining HER-2sv polypeptide is via purification from biological samples such as source tissues and/or fluids in which the HER-2sv polypeptide is naturally found. Such purification can be conducted using methods for protein purification as described herein. The presence of the HER-2sv polypeptide during purification may be monitored, for example, using an antibody prepared against recombinantly produced HER-2sv polypeptide or peptide fragments thereof.

A number of additional methods for producing nucleic acids and polypeptides are known in the art, and the methods can be used to produce polypeptides having specificity for HER-2sv polypeptide. *See, e.g.*, Roberts *et al.*, 1997, *Proc. Natl. Acad. Sci. U.S.A.* 94:12297-303, which describes the production of fusion proteins between an mRNA and its encoded peptide. *See also,* Roberts, 1999, *Curr. Opin. Chem. Biol.* 3:268-73. Additionally, U.S. Patent No. 5,824,469 describes methods for obtaining oligonucleotides capable of carrying out a specific biological function. The procedure involves generating a heterogeneous pool of oligonucleotides, each having a 5' randomized sequence, a central preselected sequence, and a 3' randomized sequence. The resulting heterogeneous pool is introduced into a population of cells that do not

exhibit the desired biological function. Subpopulations of the cells are then screened for those that exhibit a predetermined biological function. From that subpopulation, oligonucleotides capable of carrying out the desired biological function are isolated.

U.S. Patent Nos. 5,763,192; 5,814,476; 5,723,323; and 5,817,483 describe processes for producing peptides or polypeptides. This is done by producing stochastic genes or fragments thereof, and then introducing these genes into host cells which produce one or more proteins encoded by the stochastic genes. The host cells are then screened to identify those clones producing peptides or polypeptides having the desired activity.

Another method for producing peptides or polypeptides is described in International Pub. No. WO99/15650, filed by Athersys, Inc. Known as "Random Activation of Gene Expression for Gene Discovery" (RAGE-GD), the process involves the activation of endogenous gene expression or over-expression of a gene by *in situ* recombination methods. For example, expression of an endogenous gene is activated or increased by integrating a regulatory sequence into the target cell that is capable of activating expression of the gene by non-homologous or illegitimate recombination. The target DNA is first subjected to radiation, and a genetic promoter inserted. The promoter eventually locates a break at the front of a gene, initiating transcription of the gene. This results in expression of the desired peptide or polypeptide.

It will be appreciated that these methods can also be used to create comprehensive HER-2sv polypeptide expression libraries, which can subsequently be used for high throughput phenotypic screening in a variety of assays, such as biochemical assays, cellular assays, and whole organism assays (*e.g.*, plant, mouse, etc.).

6. Synthesis

It will be appreciated by those skilled in the art that the nucleic acid and polypeptide molecules described herein may be produced by recombinant and other means.

7. Selective Binding Agents

The term "selective binding agent" refers to a molecule that has specificity for one or more HER-2sv polypeptides. Suitable selective binding agents include, but are

not limited to, antibodies and derivatives thereof, polypeptides, and small molecules. Suitable selective binding agents may be prepared using methods known in the art. An exemplary HER-2SV polypeptide selective binding agent of the present invention is capable of binding a certain portion of the HER-2SV polypeptide thereby inhibiting the binding of the polypeptide to a HER-2sv polypeptide receptor.

Selective binding agents such as antibodies and antibody fragments that bind HER-2sv polypeptides are within the scope of the present invention. The antibodies may be polyclonal including monospecific polyclonal; monoclonal (MAbs); recombinant; chimeric; humanized, such as complementarity-determining region (CDR)-grafted; human; single chain; and/or bispecific; as well as fragments; variants; or derivatives thereof. Antibody fragments include those portions of the antibody that bind to an epitope on the HER-2SV polypeptide. Examples of such fragments include Fab and F(ab') fragments generated by enzymatic cleavage of full-length antibodies. Other binding fragments include those generated by recombinant DNA techniques, such as the expression of recombinant plasmids containing nucleic acid sequences encoding antibody variable regions.

Polyclonal antibodies directed toward a HER-2sv polypeptide generally are produced in animals (*e.g.*, rabbits or mice) by means of multiple subcutaneous or intraperitoneal injections of HER-2sv polypeptide and an adjuvant. It may be useful to conjugate a HER-2sv polypeptide to a carrier protein that is immunogenic in the species to be immunized, such as keyhole limpet hemocyanin, serum, albumin, bovine thyroglobulin, or soybean trypsin inhibitor. Also, aggregating agents such as alum are used to enhance the immune response. After immunization, the animals are bled and the serum is assayed for anti-HER-2sv antibody titer.

Monoclonal antibodies directed toward HER-2sv polypeptides are produced using any method that provides for the production of antibody molecules by continuous cell lines in culture. Examples of suitable methods for preparing monoclonal antibodies include the hybridoma methods of Kohler *et al.*, 1975, *Nature* 256:495-97 and the human B-cell hybridoma method (Kozbor, 1984, *J. Immunol.* 133:3001; Brodeur *et al.*, *Monoclonal Antibody Production Techniques and Applications* 51-63 (Marcel Dekker, Inc., 1987). Also provided by the invention are hybridoma cell lines that produce monoclonal antibodies reactive with HER-2sv polypeptides.

Monoclonal antibodies of the invention may be modified for use as therapeutics. One embodiment is a "chimeric" antibody in which a portion of the heavy (H) and/or light (L) chain is identical with or homologous to a corresponding sequence in antibodies derived from a particular species or belonging to a particular antibody class or subclass, while the remainder of the chain(s) is/are identical with or homologous to a corresponding sequence in antibodies derived from another species or belonging to another antibody class or subclass. Also included are fragments of such antibodies, so long as they exhibit the desired biological activity. *See* U.S. Patent No. 4,816,567; Morrison *et al.*, 1985, *Proc. Natl. Acad. Sci.* 81:6851-55.

In another embodiment, a monoclonal antibody of the invention is a "humanized" antibody. Methods for humanizing non-human antibodies are well known in the art. *See* U.S. Patent Nos. 5,585,089 and 5,693,762. Generally, a humanized antibody has one or more amino acid residues introduced into it from a source that is non-human. Humanization can be performed, for example, using methods described in the art (Jones *et al.*, 1986, *Nature* 321:522-25; Riechmann *et al.*, 1998, *Nature* 332:323-27; Verhoeyen *et al.*, 1988, *Science* 239:1534-36), by substituting at least a portion of a rodent complementarity-determining region for the corresponding regions of a human antibody.

Also encompassed by the invention are human antibodies that bind HER-2sv polypeptides. Using transgenic animals (*e.g.*, mice) that are capable of producing a repertoire of human antibodies in the absence of endogenous immunoglobulin production such antibodies are produced by immunization with a HER-2sv polypeptide antigen (*i.e.*, having at least 6 contiguous amino acids), optionally conjugated to a carrier. *See, e.g.*, Jakobovits *et al.*, 1993, *Proc. Natl. Acad. Sci.* 90:2551-55; Jakobovits *et al.*, 1993, *Nature* 362:255-58; Bruggermann *et al.*, 1993, *Year in Immuno.* 7:33. In one method, such transgenic animals are produced by incapacitating the endogenous loci encoding the heavy and light immunoglobulin chains therein, and inserting loci encoding human heavy and light chain proteins into the genome thereof. Partially modified animals (*i.e.*, those having less than the full complement of modifications) are then cross-bred to obtain an animal having all of the desired immune system modifications. When administered an immunogen, these transgenic animals produce antibodies with human (rather than, *e.g.*, murine) amino acid sequences, including variable regions that are immunospecific for these antigens. *See* International App. Nos. PCT/US96/05928 and PCT/US93/06926. Additional

methods are described in U.S. Patent No. 5,545,807, International App. Nos. PCT/US91/245 and PCT/GB89/01207, and in European Patent Nos. 546073B1 and 546073A1. Human antibodies can also be produced by the expression of recombinant DNA in host cells or by expression in hybridoma cells as described herein.

In an alternative embodiment, human antibodies can also be produced from phage-display libraries (Hoogenboom *et al.*, 1991, *J. Mol. Biol.* 227:381; Marks *et al.*, 1991, *J. Mol. Biol.* 222:581). These processes mimic immune selection through the display of antibody repertoires on the surface of filamentous bacteriophage, and subsequent selection of phage by their binding to an antigen of choice. One such technique is described in International App. No. PCT/US98/17364, which describes the isolation of high affinity and functional agonistic antibodies for MPL- and msk-receptors using such an approach.

Chimeric, CDR grafted, and humanized antibodies are typically produced by recombinant methods. Nucleic acids encoding the antibodies are introduced into host cells and expressed using materials and procedures described herein. In a preferred embodiment, the antibodies are produced in mammalian host cells, such as CHO cells. Monoclonal (*e.g.*, human) antibodies may be produced by the expression of recombinant DNA in host cells or by expression in hybridoma cells as described herein.

The anti-HER-2sv antibodies of the invention may be employed in any known assay method, such as competitive binding assays, direct and indirect sandwich assays, and immunoprecipitation assays (Sola, *Monoclonal Antibodies: A Manual of Techniques* 147-158 (CRC Press, Inc., 1987)) for the detection and quantitation of HER-2sv polypeptides. The antibodies will bind HER-2sv polypeptides with an affinity that is appropriate for the assay method being employed.

For diagnostic applications, in certain embodiments, anti-HER-2sv antibodies may be labeled with a detectable moiety. The detectable moiety can be any one that is capable of producing, either directly or indirectly, a detectable signal. For example, the detectable moiety may be a radioisotope, such as ^{3}H, ^{14}C, ^{32}P, ^{35}S, ^{125}I, ^{99}Tc, ^{111}In, or ^{67}Ga; a fluorescent or chemiluminescent compound, such as fluorescein isothiocyanate, rhodamine, or luciferin; or an enzyme, such as alkaline phosphatase, β-galactosidase, or horseradish peroxidase (Bayer, *et al.*, 1990, *Meth. Enz.* 184:138-63).

Competitive binding assays rely on the ability of a labeled standard (*e.g.*, a HER-2sv polypeptide, or an immunologically reactive portion thereof) to compete with the test sample analyte (an HER-2sv polypeptide) for binding with a limited amount of anti-HER-2sv antibody. The amount of a HER-2sv polypeptide in the test sample is inversely proportional to the amount of standard that becomes bound to the antibodies. To facilitate determining the amount of standard that becomes bound, the antibodies typically are insolubilized before or after the competition, so that the standard and analyte that are bound to the antibodies may conveniently be separated from the standard and analyte that remain unbound.

Sandwich assays typically involve the use of two antibodies, each capable of binding to a different immunogenic portion, or epitope, of the protein to be detected and/or quantitated. In a sandwich assay, the test sample analyte is typically bound by a first antibody that is immobilized on a solid support, and thereafter a second antibody binds to the analyte, thus forming an insoluble three-part complex. *See, e.g.*, U.S. Patent No. 4,376,110. The second antibody may itself be labeled with a detectable moiety (direct sandwich assays) or may be measured using an anti-immunoglobulin antibody that is labeled with a detectable moiety (indirect sandwich assays). For example, one type of sandwich assay is an enzyme-linked immunosorbent assay (ELISA), in which case the detectable moiety is an enzyme.

The selective binding agents, including anti-HER-2sv antibodies, are also useful for *in vivo* imaging. An antibody labeled with a detectable moiety may be administered to an animal, preferably into the bloodstream, and the presence and location of the labeled antibody in the host assayed. The antibody may be labeled with any moiety that is detectable in an animal, whether by nuclear magnetic resonance, radiology, or other detection means known in the art.

Selective binding agents of the invention, including antibodies, may be used as therapeutics. In a preferred embodiment, the selective binding agent is an antagonist antibody capable of specifically binding to a HER-2sv polypeptide thereby inhibiting or eliminating the functional activity of a HER-2sv polypeptide *in vivo* or *in vitro*. In preferred embodiments, the selective binding agent, *e.g.*, an antagonist antibody, will inhibit the functional activity of a HER-2sv polypeptide by at least about 50%, and preferably by at least about 80%. In another embodiment, the selective binding agent may be an anti-HER-2sv polypeptide antibody that is capable of interacting with a HER-2sv polypeptide binding partner (a ligand or receptor) thereby inhibiting or

eliminating HER-2sv polypeptide activity *in vitro* or *in vivo*. Selective binding agents, including antagonist anti-HER-2sv polypeptide antibodies, are identified by screening assays that are well known in the art.

The invention also relates to a kit comprising HER-2sv selective binding agents (such as antibodies) and other reagents useful for detecting HER-2sv polypeptide levels in biological samples. Such reagents may include a detectable label, blocking serum, positive and negative control samples, and detection reagents.

8. Microarrays

It will be appreciated that DNA microarray technology can be utilized in accordance with the present invention. DNA microarrays are miniature, high-density arrays of nucleic acids positioned on a solid support, such as glass. Each cell or element within the array contains numerous copies of a single nucleic acid species that acts as a target for hybridization with a complementary nucleic acid sequence (*e.g.*, mRNA). In expression profiling using DNA microarray technology, mRNA is first extracted from a cell or tissue sample and then converted enzymatically to fluorescently labeled cDNA. This material is hybridized to the microarray and unbound cDNA is removed by washing. The expression of discrete genes represented on the array is then visualized by quantitating the amount of labeled cDNA that is specifically bound to each target nucleic acid molecule. In this way, the expression of thousands of genes can be quantitated in a high throughput, parallel manner from a single sample of biological material.

This high throughput expression profiling has a broad range of applications with respect to the HER-2sv molecules of the invention, including, but not limited to: the identification and validation of HER-2sv disease-related genes as targets for therapeutics; molecular toxicology of related HER-2sv molecules and inhibitors thereof; stratification of populations and generation of surrogate markers for clinical trials; and enhancing related HER-2sv polypeptide small molecule drug discovery by aiding in the identification of selective compounds in high throughput screens.

9. Chemical Derivatives

Chemically modified derivatives of HER-2sv polypeptides may be prepared by one skilled in the art, given the disclosures described herein. HER-2sv polypeptide derivatives are modified in a manner that is different – either in the type or location of

the molecules naturally attached to the polypeptide. Derivatives may include molecules formed by the deletion of one or more naturally-attached chemical groups. The polypeptide comprising the amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, or other HER-2sv polypeptide, may be modified by the covalent attachment of one or more polymers. For example, the polymer selected is typically water-soluble so that the protein to which it is attached does not precipitate in an aqueous environment, such as a physiological environment. Included within the scope of suitable polymers is a mixture of polymers. Preferably, for therapeutic use of the end-product preparation, the polymer will be pharmaceutically acceptable.

The polymers each may be of any molecular weight and may be branched or unbranched. The polymers each typically have an average molecular weight of between about 2 kDa to about 100 kDa (the term "about" indicating that in preparations of a water-soluble polymer, some molecules will weigh more, some less, than the stated molecular weight). The average molecular weight of each polymer is preferably between about 5 kDa and about 50 kDa, more preferably between about 12 kDa and about 40 kDa and most preferably between about 20 kDa and about 35 kDa.

Suitable water-soluble polymers or mixtures thereof include, but are not limited to, N-linked or O-linked carbohydrates, sugars, phosphates, polyethylene glycol (PEG) (including the forms of PEG that have been used to derivatize proteins, including mono-(C_1-C_{10}), alkoxy-, or aryloxy-polyethylene glycol), monomethoxy-polyethylene glycol, dextran (such as low molecular weight dextran of, for example, about 6 kD), cellulose, or other carbohydrate based polymers, poly-(N-vinyl pyrrolidone) polyethylene glycol, propylene glycol homopolymers, polypropylene oxide/ethylene oxide co-polymers, polyoxyethylated polyols (*e.g.*, glycerol), and polyvinyl alcohol. Also encompassed by the present invention are bifunctional crosslinking molecules that may be used to prepare covalently attached HER-2sv polypeptide multimers.

In general, chemical derivatization may be performed under any suitable condition used to react a protein with an activated polymer molecule. Methods for preparing chemical derivatives of polypeptides will generally comprise the steps of: (a) reacting the polypeptide with the activated polymer molecule (such as a reactive ester or aldehyde derivative of the polymer molecule) under conditions whereby the polypeptide comprising the amino acid sequence of any of SEQ ID NO: 2, SEQ ID

NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, or other HER-2sv polypeptide, becomes attached to one or more polymer molecules, and (b) obtaining the reaction products. The optimal reaction conditions will be determined based on known parameters and the desired result. For example, the larger the ratio of polymer molecules to protein, the greater the percentage of attached polymer molecule. In one embodiment, the HER-2sv polypeptide derivative may have a single polymer molecule moiety at the amino-terminus. *See, e.g.*, U.S. Patent No. 5,234,784.

The pegylation of a polypeptide may be specifically carried out using any of the pegylation reactions known in the art. Such reactions are described, for example, in the following references: Francis *et al.*, 1992, *Focus on Growth Factors* 3:4-10; European Patent Nos. 0154316 and 0401384; and U.S. Patent No. 4,179,337. For example, pegylation may be carried out via an acylation reaction or an alkylation reaction with a reactive polyethylene glycol molecule (or an analogous reactive water-soluble polymer) as described herein. For the acylation reactions, a selected polymer should have a single reactive ester group. For reductive alkylation, a selected polymer should have a single reactive aldehyde group. A reactive aldehyde is, for example, polyethylene glycol propionaldehyde, which is water stable, or mono C_1-C_{10} alkoxy or aryloxy derivatives thereof (*see* U.S. Patent No. 5,252,714).

In another embodiment, HER-2sv polypeptides may be chemically coupled to biotin. The biotin/HER-2sv polypeptide molecules are then allowed to bind to avidin, resulting in tetravalent avidin/biotin/HER-2sv polypeptide molecules. HER-2sv polypeptides may also be covalently coupled to dinitrophenol (DNP) or trinitrophenol (TNP) and the resulting conjugates precipitated with anti-DNP or anti-TNP-IgM to form decameric conjugates with a valency of 10.

Generally, conditions that may be alleviated or modulated by the administration of the present HER-2sv polypeptide derivatives include those described herein for HER-2sv polypeptides. However, the HER-2sv polypeptide derivatives disclosed herein may have additional activities, enhanced or reduced biological activity, or other characteristics, such as increased or decreased half-life, as compared to the non-derivatized molecules.

10. Genetically Engineered Non-Human Animals

Additionally included within the scope of the present invention are non-human animals such as mice, rats, or other rodents; rabbits, goats, sheep, or other

farm animals, in which the genes encoding native HER-2sv polypeptide have been disrupted (*i.e.*, "knocked out") such that the level of expression of HER-2sv polypeptide is significantly decreased or completely abolished. Such animals may be prepared using techniques and methods such as those described in U.S. Patent No. 5,557,032.

The present invention further includes non-human animals such as mice, rats, or other rodents; rabbits, goats, sheep, or other farm animals, in which either the native form of a HER-2sv gene for that animal or a heterologous HER-2sv gene is over-expressed by the animal, thereby creating a "transgenic" animal. Such transgenic animals may be prepared using well known methods such as those described in U.S. Patent No 5,489,743 and International Pub. No. WO 94/28122.

The present invention further includes non-human animals in which the promoter for one or more of the HER-2sv polypeptides of the present invention is either activated or inactivated (*e.g.*, by using homologous recombination methods) to alter the level of expression of one or more of the native HER-2sv polypeptides.

These non-human animals may be used for drug candidate screening. In such screening, the impact of a drug candidate on the animal may be measured. For example, drug candidates may decrease or increase the expression of the HER-2sv gene. In certain embodiments, the amount of HER-2sv polypeptide that is produced may be measured after the exposure of the animal to the drug candidate. Additionally, in certain embodiments, one may detect the actual impact of the drug candidate on the animal. For example, over-expression of a particular gene may result in, or be associated with, a disease or pathological condition. In such cases, one may test a drug candidate's ability to decrease expression of the gene or its ability to prevent or inhibit a pathological condition. In other examples, the production of a particular metabolic product such as a fragment of a polypeptide, may result in, or be associated with, a disease or pathological condition. In such cases, one may test a drug candidate's ability to decrease the production of such a metabolic product or its ability to prevent or inhibit a pathological condition.

11. Assaying for Other Modulators of HER-2sv Polypeptide Activity

In some situations, it may be desirable to identify molecules that are modulators, *e.g.*, antagonists, of the activity of HER-2sv polypeptide. Natural or synthetic molecules that modulate HER-2sv polypeptide may be identified using one

or more screening assays, such as those described herein. Such molecules may be administered either in an *ex vivo* manner or in an *in vivo* manner by injection, or by oral delivery, implantation device, or the like.

"Test molecule" refers to a molecule that is under evaluation for the ability to modulate (*i.e.*, increase or decrease) the activity of a HER-2sv polypeptide. Most commonly, a test molecule will interact directly with a HER-2sv polypeptide. However, it is also contemplated that a test molecule may also modulate HER-2sv polypeptide activity indirectly, such as by affecting HER-2sv gene expression, or by binding to a HER-2sv polypeptide binding partner (*e.g.*, receptor or ligand). In one embodiment, a test molecule will bind to a HER-2sv polypeptide with an affinity constant of at least about 10^{-6} M, preferably about 10^{-8} M, more preferably about 10^{-9} M, and even more preferably about 10^{-10} M.

Methods for identifying compounds that interact with HER-2sv polypeptides are encompassed by the present invention. In certain embodiments, a HER-2sv polypeptide is incubated with a test molecule under conditions that permit the interaction of the test molecule with a HER-2sv polypeptide, and the extent of the interaction is measured. The test molecule can be screened in a substantially purified form or in a crude mixture.

In certain embodiments, a HER-2sv polypeptide antagonist may be a protein, peptide, carbohydrate, lipid, or small molecular weight molecule that interacts with HER-2sv polypeptide to regulate its activity. Molecules which regulate HER-2sv polypeptide expression include nucleic acids which are complementary to nucleic acids encoding a HER-2sv polypeptide, or are complementary to nucleic acids sequences which direct or control the expression of HER-2sv polypeptide, and which act as anti-sense regulators of expression.

Once a test molecule has been identified as interacting with a HER-2sv polypeptide, the molecule may be further evaluated for its ability to increase or decrease HER-2sv polypeptide activity. The measurement of the interaction of a test molecule with HER-2sv polypeptide may be carried out in several formats, including cell-based binding assays, membrane binding assays, solution-phase assays, and immunoassays. In general, a test molecule is incubated with a HER-2sv polypeptide for a specified period of time, and HER-2sv polypeptide activity is determined by one or more assays for measuring biological activity.

The interaction of test molecules with HER-2sv polypeptides may also be assayed directly using polyclonal or monoclonal antibodies in an immunoassay. Alternatively, modified forms of HER-2sv polypeptides containing epitope tags as described herein may be used in solution and immunoassays.

In the event that HER-2sv polypeptides display biological activity through an interaction with a binding partner (*e.g.*, a receptor or a ligand), a variety of *in vitro* assays may be used to measure the binding of a HER-2sv polypeptide to the corresponding binding partner (such as a selective binding agent, receptor, or ligand). These assays may be used to screen test molecules for their ability to increase or decrease the rate and/or the extent of binding of a HER-2sv polypeptide to its binding partner. In one assay, a HER-2sv polypeptide is immobilized in the wells of a microtiter plate. Radiolabeled HER-2sv polypeptide binding partner (for example, iodinated HER-2sv polypeptide binding partner) and a test molecule can then be added either one at a time (in either order) or simultaneously to the wells. After incubation, the wells can be washed and counted for radioactivity, using a scintillation counter, to determine the extent to which the binding partner bound to the HER-2sv polypeptide. Typically, a molecule will be tested over a range of concentrations, and a series of control wells lacking one or more elements of the test assays can be used for accuracy in the evaluation of the results. An alternative to this method involves reversing the "positions" of the proteins, *i.e.*, immobilizing HER-2sv polypeptide binding partner to the microtiter plate wells, incubating with the test molecule and radiolabeled HER-2sv polypeptide, and determining the extent of HER-2sv polypeptide binding. *See, e.g., Current Protocols in Molecular Biology*, chap. 18 (Ausubel *et al.*, eds., Green Publishers Inc. and Wiley and Sons 1995).

As an alternative to radiolabeling, a HER-2sv polypeptide or its binding partner may be conjugated to biotin, and the presence of biotinylated protein can then be detected using streptavidin linked to an enzyme, such as horse radish peroxidase (HRP) or alkaline phosphatase (AP), which can be detected colorometrically, or by fluorescent tagging of streptavidin. An antibody directed to a HER-2sv polypeptide or to a HER-2sv polypeptide binding partner, and which is conjugated to biotin, may also be used for purposes of detection following incubation of the complex with enzyme-linked streptavidin linked to AP or HRP.

A HER-2sv polypeptide or a HER-2sv polypeptide binding partner can also be immobilized by attachment to agarose beads, acrylic beads, or other types of such

inert solid phase substrates. The substrate-protein complex can be placed in a solution containing the complementary protein and the test compound. After incubation, the beads can be precipitated by centrifugation, and the amount of binding between a HER-2sv polypeptide and its binding partner can be assessed using the methods described herein. Alternatively, the substrate-protein complex can be immobilized in a column with the test molecule and complementary protein passing through the column. The formation of a complex between a HER-2sv polypeptide and its binding partner can then be assessed using any of the techniques described herein (*e.g.*, radiolabelling or antibody binding).

Another *in vitro* assay that is useful for identifying a test molecule that increases or decreases the formation of a complex between a HER-2sv polypeptide binding protein and a HER-2sv polypeptide binding partner is a surface plasmon resonance detector system such as the BIAcore assay system (Pharmacia, Piscataway, NJ). The BIAcore system is utilized as specified by the manufacturer. This assay essentially involves the covalent binding of either HER-2sv polypeptide or a HER-2sv polypeptide binding partner to a dextran-coated sensor chip that is located in a detector. The test compound and the other complementary protein can then be injected, either simultaneously or sequentially, into the chamber containing the sensor chip. The amount of complementary protein that binds can be assessed based on the change in molecular mass that is physically associated with the dextran-coated side of the sensor chip, with the change in molecular mass being measured by the detector system.

In some cases, it may be desirable to evaluate two or more test compounds together for their ability to increase or decrease the formation of a complex between a HER-2sv polypeptide and a HER-2sv polypeptide binding partner. In these cases, the assays set forth herein can be readily modified by adding such additional test compound(s) either simultaneously with, or subsequent to, the first test compound. The remainder of the steps in the assay are as set forth herein.

In vitro assays such as those described herein may be used advantageously to screen large numbers of compounds for an effect on the formation of a complex between a HER-2sv polypeptide and HER-2sv polypeptide binding partner. The assays may be automated to screen compounds generated in phage display, synthetic peptide, and chemical synthesis libraries.

Compounds which increase or decrease the formation of a complex between a HER-2sv polypeptide and a HER-2sv polypeptide binding partner may also be screened in cell culture using cells and cell lines expressing either HER-2sv polypeptide or HER-2sv polypeptide binding partner. Cells and cell lines may be obtained from any mammal, but preferably will be from human or other primate, canine, or rodent sources. The binding of a HER-2sv polypeptide to cells expressing HER-2sv polypeptide binding partner at the surface is evaluated in the presence or absence of test molecules, and the extent of binding may be determined by, for example, flow cytometry using a biotinylated antibody to a HER-2sv polypeptide binding partner. Cell culture assays can be used advantageously to further evaluate compounds that score positive in protein binding assays described herein.

Cell cultures can also be used to screen the impact of a drug candidate. For example, drug candidates may decrease or increase the expression of the HER-2sv gene. In certain embodiments, the amount of HER-2sv polypeptide or a HER-2sv polypeptide fragment that is produced may be measured after exposure of the cell culture to the drug candidate. In certain embodiments, one may detect the actual impact of the drug candidate on the cell culture. For example, the over-expression of a particular gene may have a particular impact on the cell culture. In such cases, one may test a drug candidate's ability to increase or decrease the expression of the gene or its ability to prevent or inhibit a particular impact on the cell culture. In other examples, the production of a particular metabolic product such as a fragment of a polypeptide, may result in, or be associated with, a disease or pathological condition. In such cases, one may test a drug candidate's ability to decrease the production of such a metabolic product in a cell culture.

12. Internalizing Proteins

The *tat* protein sequence (from HIV) can be used to internalize proteins into a cell. *See, e.g.*, Falwell *et al.*, 1994, *Proc. Natl. Acad. Sci. U.S.A.* 91:664-68. For example, an 11 amino acid sequence (Y-G-R-K-K-R-R-Q-R-R-R; SEQ ID NO: 12) of the HIV tat protein (termed the "protein transduction domain," or TAT PDT) has been described as mediating delivery across the cytoplasmic membrane and the nuclear membrane of a cell. *See* Schwarze *et al.*, 1999, *Science* 285:1569-72; and Nagahara *et al.*, 1998, *Nat. Med.* 4:1449-52. In these procedures, FITC-constructs (FITC-labeled G-G-G-G-Y-G-R-K-K-R-R-Q-R-R-R; SEQ ID NO: 13), which penetrate

tissues following intraperitoneal administration, are prepared, and the binding of such constructs to cells is detected by fluorescence-activated cell sorting (FACS) analysis. Cells treated with a *tat*-β-gal fusion protein will demonstrate β-gal activity. Following injection, expression of such a construct can be detected in a number of tissues, including liver, kidney, lung, heart, and brain tissue. It is believed that such constructs undergo some degree of unfolding in order to enter the cell, and as such, may require a refolding following entry into the cell.

It will thus be appreciated that the *tat* protein sequence may be used to internalize a desired polypeptide into a cell. For example, using the *tat* protein sequence, a HER-2sv antagonist (such as an anti-HER-2sv selective binding agent, small molecule, soluble receptor, or antisense oligonucleotide) can be administered intracellularly to inhibit the activity of a HER-2sv molecule. As used herein, the term "HER-2sv molecule" refers to both HER-2sv nucleic acid molecules and HER-2sv polypeptides as defined herein. Where desired, the HER-2sv protein itself may also be internally administered to a cell using these procedures. *See also*, Straus, 1999, *Science* 285:1466-67.

13. Cell Source Identification Using HER-2sv Polypeptide

In accordance with certain embodiments of the invention, it may be useful to be able to determine the source of a certain cell type associated with a HER-2sv polypeptide. For example, it may be useful to determine the origin of a disease or pathological condition as an aid in selecting an appropriate therapy. In certain embodiments, nucleic acids encoding a HER-2sv polypeptide can be used as a probe to identify cells described herein by screening the nucleic acids of the cells with such a probe. In other embodiments, one may use anti-HER-2sv polypeptide antibodies to test for the presence of HER-2sv polypeptide in cells, and thus, determine if such cells are of the types described herein.

14. HER-2sv Polypeptide Compositions and Administration

Therapeutic compositions are within the scope of the present invention. Such HER-2SV polypeptide pharmaceutical compositions may comprise a therapeutically effective amount of a HER-2sv polypeptide or a HER-2sv nucleic acid molecule in admixture with a pharmaceutically or physiologically acceptable formulation agent selected for suitability with the mode of administration. Pharmaceutical compositions

may comprise a therapeutically effective amount of one or more HER-2sv polypeptide selective binding agents in admixture with a pharmaceutically or physiologically acceptable formulation agent selected for suitability with the mode of administration.

Acceptable formulation materials preferably are nontoxic to recipients at the dosages and concentrations employed.

The pharmaceutical composition may contain formulation materials for modifying, maintaining, or preserving, for example, the pH, osmolarity, viscosity, clarity, color, isotonicity, odor, sterility, stability, rate of dissolution or release, adsorption, or penetration of the composition. Suitable formulation materials include, but are not limited to, amino acids (such as glycine, glutamine, asparagine, arginine, or lysine), antimicrobials, antioxidants (such as ascorbic acid, sodium sulfite, or sodium hydrogen-sulfite), buffers (such as borate, bicarbonate, Tris-HCl, citrates, phosphates, or other organic acids), bulking agents (such as mannitol or glycine), chelating agents (such as ethylenediamine tetraacetic acid (EDTA)), complexing agents (such as caffeine, polyvinylpyrrolidone, beta-cyclodextrin, or hydroxypropyl-beta-cyclodextrin), fillers, monosaccharides, disaccharides, and other carbohydrates (such as glucose, mannose, or dextrins), proteins (such as serum albumin, gelatin, or immunoglobulins), coloring, flavoring and diluting agents, emulsifying agents, hydrophilic polymers (such as polyvinylpyrrolidone), low molecular weight polypeptides, salt-forming counterions (such as sodium), preservatives (such as benzalkonium chloride, benzoic acid, salicylic acid, thimerosal, phenethyl alcohol, methylparaben, propylparaben, chlorhexidine, sorbic acid, or hydrogen peroxide), solvents (such as glycerin, propylene glycol, or polyethylene glycol), sugar alcohols (such as mannitol or sorbitol), suspending agents, surfactants or wetting agents (such as pluronics; PEG; sorbitan esters; polysorbates such as polysorbate 20 or polysorbate 80; triton; tromethamine; lecithin; cholesterol or tyloxapal), stability enhancing agents (such as sucrose or sorbitol), tonicity enhancing agents (such as alkali metal halides – preferably sodium or potassium chloride – or mannitol sorbitol), delivery vehicles, diluents, excipients and/or pharmaceutical adjuvants. *See Remington's Pharmaceutical Sciences* (18th Ed., A.R. Gennaro, ed., Mack Publishing Company 1990.

The optimal pharmaceutical composition will be determined by a skilled artisan depending upon, for example, the intended route of administration, delivery

format, and desired dosage. *See, e.g., Remington's Pharmaceutical Sciences, supra.* Such compositions may influence the physical state, stability, rate of *in vivo* release, and rate of *in vivo* clearance of the HER-2sv molecule.

The primary vehicle or carrier in a pharmaceutical composition may be either aqueous or non-aqueous in nature. For example, a suitable vehicle or carrier for injection may be water, physiological saline solution, or artificial cerebrospinal fluid, possibly supplemented with other materials common in compositions for parenteral administration. Neutral buffered saline or saline mixed with serum albumin are further exemplary vehicles. Other exemplary pharmaceutical compositions comprise Tris buffer of about pH 7.0-8.5, or acetate buffer of about pH 4.0-5.5, which may further include sorbitol or a suitable substitute. In one embodiment of the present invention, HER-2sv polypeptide compositions may be prepared for storage by mixing the selected composition having the desired degree of purity with optional formulation agents (*Remington's Pharmaceutical Sciences, supra*) in the form of a lyophilized cake or an aqueous solution. Further, the HER-2sv polypeptide product may be formulated as a lyophilizate using appropriate excipients such as sucrose.

The HER-2sv polypeptide pharmaceutical compositions can be selected for parenteral delivery. Alternatively, the compositions may be selected for inhalation or for delivery through the digestive tract, such as orally. The preparation of such pharmaceutically acceptable compositions is within the skill of the art.

The formulation components are present in concentrations that are acceptable to the site of administration. For example, buffers are used to maintain the composition at physiological pH or at a slightly lower pH, typically within a pH range of from about 5 to about 8.

When parenteral administration is contemplated, the therapeutic compositions for use in this invention may be in the form of a pyrogen-free, parenterally acceptable, aqueous solution comprising the desired HER-2sv molecule in a pharmaceutically acceptable vehicle. A particularly suitable vehicle for parenteral injection is sterile distilled water in which a HER-2sv molecule is formulated as a sterile, isotonic solution, properly preserved. Yet another preparation can involve the formulation of the desired molecule with an agent, such as injectable microspheres, bio-erodible particles, polymeric compounds (such as polylactic acid or polyglycolic acid), beads, or liposomes, that provides for the controlled or sustained release of the product which may then be delivered via a depot injection. Hyaluronic acid may also be used,

and this may have the effect of promoting sustained duration in the circulation. Other suitable means for the introduction of the desired molecule include implantable drug delivery devices.

In one embodiment, a pharmaceutical composition may be formulated for inhalation. For example, HER-2sv polypeptide may be formulated as a dry powder for inhalation. HER-2sv polypeptide or nucleic acid molecule inhalation solutions may also be formulated with a propellant for aerosol delivery. In yet another embodiment, solutions may be nebulized. Pulmonary administration is further described in International Pub. No. WO 94/20069, which describes the pulmonary delivery of chemically modified proteins.

It is also contemplated that certain formulations may be administered orally. In one embodiment of the present invention, HER-2sv polypeptides that are administered in this fashion can be formulated with or without those carriers customarily used in the compounding of solid dosage forms such as tablets and capsules. For example, a capsule may be designed to release the active portion of the formulation at the point in the gastrointestinal tract when bioavailability is maximized and pre-systemic degradation is minimized. Additional agents can be included to facilitate absorption of the HER-2sv polypeptide. Diluents, flavorings, low melting point waxes, vegetable oils, lubricants, suspending agents, tablet disintegrating agents, and binders may also be employed.

Another pharmaceutical composition may involve an effective quantity of HER-2sv polypeptides in a mixture with non-toxic excipients that are suitable for the manufacture of tablets. By dissolving the tablets in sterile water, or another appropriate vehicle, solutions can be prepared in unit-dose form. Suitable excipients include, but are not limited to, inert diluents, such as calcium carbonate, sodium carbonate or bicarbonate, lactose, or calcium phosphate; or binding agents, such as starch, gelatin, or acacia; or lubricating agents such as magnesium stearate, stearic acid, or talc.

Additional HER-2sv polypeptide pharmaceutical compositions will be evident to those skilled in the art, including formulations involving HER-2sv polypeptides in sustained- or controlled-delivery formulations. Techniques for formulating a variety of other sustained- or controlled-delivery means, such as liposome carriers, bio-erodible microparticles or porous beads and depot injections, are also known to those skilled in the art. *See, e.g.*, International App. No. PCT/US93/00829, which describes

the controlled release of porous polymeric microparticles for the delivery of pharmaceutical compositions.

Additional examples of sustained-release preparations include semipermeable polymer matrices in the form of shaped articles, *e.g.* films, or microcapsules. Sustained release matrices may include polyesters, hydrogels, polylactides (U.S. Patent No. 3,773,919 and European Patent No. 058481), copolymers of L-glutamic acid and gamma ethyl-L-glutamate (Sidman *et al.*, 1983, *Biopolymers* 22:547-56), poly(2-hydroxyethyl-methacrylate) (Langer *et al.*, 1981, *J. Biomed. Mater. Res.* 15:167-277 and Langer, 1982, *Chem. Tech.* 12:98-105), ethylene vinyl acetate (Langer *et al.*, *supra*) or poly-D(-)-3-hydroxybutyric acid (European Patent No. 133988). Sustained-release compositions may also include liposomes, which can be prepared by any of several methods known in the art. *See, e.g.*, Eppstein *et al.*, 1985, *Proc. Natl. Acad. Sci. USA* 82:3688-92; and European Patent Nos. 036676, 088046, and 143949.

The HER-2sv pharmaceutical composition to be used for *in vivo* administration typically must be sterile. This may be accomplished by filtration through sterile filtration membranes. Where the composition is lyophilized, sterilization using this method may be conducted either prior to, or following, lyophilization and reconstitution. The composition for parenteral administration may be stored in lyophilized form or in a solution. In addition, parenteral compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Once the pharmaceutical composition has been formulated, it may be stored in sterile vials as a solution, suspension, gel, emulsion, solid, or as a dehydrated or lyophilized powder. Such formulations may be stored either in a ready-to-use form or in a form (*e.g.*, lyophilized) requiring reconstitution prior to administration.

In a specific embodiment, the present invention is directed to kits for producing a single-dose administration unit. The kits may each contain both a first container having a dried protein and a second container having an aqueous formulation. Also included within the scope of this invention are kits containing single and multi-chambered pre-filled syringes (*e.g.*, liquid syringes and lyosyringes).

The effective amount of a HER-2sv pharmaceutical composition to be employed therapeutically will depend, for example, upon the therapeutic context and

objectives. One skilled in the art will appreciate that the appropriate dosage levels for treatment will thus vary depending, in part, upon the molecule delivered, the indication for which the HER-2sv molecule is being used, the route of administration, and the size (body weight, body surface, or organ size) and condition (the age and general health) of the patient. Accordingly, the clinician may titer the dosage and modify the route of administration to obtain the optimal therapeutic effect. A typical dosage may range from about 0.1 μg/kg to up to about 100 mg/kg or more, depending on the factors mentioned above. In other embodiments, the dosage may range from 0.1 μg/kg up to about 100 mg/kg; or 1 μg/kg up to about 100 mg/kg; or 5 μg/kg up to about 100 mg/kg.

The frequency of dosing will depend upon the pharmacokinetic parameters of the HER-2sv molecule in the formulation being used. Typically, a clinician will administer the composition until a dosage is reached that achieves the desired effect. The composition may therefore be administered as a single dose, as two or more doses (which may or may not contain the same amount of the desired molecule) over time, or as a continuous infusion via an implantation device or catheter. Further refinement of the appropriate dosage is routinely made by those of ordinary skill in the art and is within the ambit of tasks routinely performed by them. Appropriate dosages may be ascertained through use of appropriate dose-response data.

The route of administration of the pharmaceutical composition is in accord with known methods, *e.g.*, orally; through injection by intravenous, intraperitoneal, intracerebral (intraparenchymal), intracerebroventricular, intramuscular, intraocular, intraarterial, intraportal, or intralesional routes; by sustained release systems; or by implantation devices. Where desired, the compositions may be administered by bolus injection or continuously by infusion, or by implantation device.

Alternatively or additionally, the composition may be administered locally via implantation of a membrane, sponge, or other appropriate material onto which the desired molecule has been absorbed or encapsulated. Where an implantation device is used, the device may be implanted into any suitable tissue or organ, and delivery of the desired molecule may be via diffusion, timed-release bolus, or continuous administration.

In some cases, it may be desirable to use HER-2sv polypeptide pharmaceutical compositions in an *ex vivo* manner. In such instances, cells, tissues, or organs that have been removed from the patient are exposed to HER-2sv

polypeptide pharmaceutical compositions after which the cells, tissues, or organs are subsequently implanted back into the patient.

In other cases, a HER-2sv polypeptide can be delivered by implanting certain cells that have been genetically engineered, using methods such as those described herein, to express and secrete the HER-2sv polypeptide. Such cells may be animal or human cells, and may be autologous, heterologous, or xenogeneic. Optionally, the cells may be immortalized. In order to decrease the chance of an immunological response, the cells may be encapsulated to avoid infiltration of surrounding tissues. The encapsulation materials are typically biocompatible, semi-permeable polymeric enclosures or membranes that allow the release of the protein product(s) but prevent the destruction of the cells by the patient's immune system or by other detrimental factors from the surrounding tissues.

As discussed herein, it may be desirable to treat isolated cell populations (such as stem cells, lymphocytes, red blood cells, chondrocytes, neurons, and the like) with one or more HER-2sv polypeptides. This can be accomplished by exposing the isolated cells to the polypeptide directly, where it is in a form that is permeable to the cell membrane.

Additional embodiments of the present invention relate to cells and methods (*e.g.*, homologous recombination and/or other recombinant production methods) for both the *in vitro* production of therapeutic polypeptides and for the production and delivery of therapeutic polypeptides by gene therapy or cell therapy. Homologous and other recombination methods may be used to modify a cell that contains a normally transcriptionally-silent HER-2sv gene, or an under-expressed gene, and thereby produce a cell which expresses therapeutically efficacious amounts of HER-2sv polypeptides.

Homologous recombination is a technique originally developed for targeting genes to induce or correct mutations in transcriptionally active genes. Kucherlapati, 1989, *Prog. in Nucl. Acid Res. & Mol. Biol.* 36:301. The basic technique was developed as a method for introducing specific mutations into specific regions of the mammalian genome (Thomas *et al.*, 1986, *Cell* 44:419-28; Thomas and Capecchi, 1987, *Cell* 51:503-12; Doetschman *et al.*, 1988, *Proc. Natl. Acad. Sci. U.S.A.* 85:8583-87) or to correct specific mutations within defective genes (Doetschman *et al.*, 1987, *Nature* 330:576-78). Exemplary homologous recombination techniques are

described in U.S. Patent No. 5,272,071; European Patent Nos. 9193051 and 505500; International App. No. PCT/US90/07642, and International Pub No. WO 91/09955).

Through homologous recombination, the DNA sequence to be inserted into the genome can be directed to a specific region of the gene of interest by attaching it to targeting DNA. The targeting DNA is a nucleotide sequence that is complementary (homologous) to a region of the genomic DNA. Small pieces of targeting DNA that are complementary to a specific region of the genome are put in contact with the parental strand during the DNA replication process. It is a general property of DNA that has been inserted into a cell to hybridize, and therefore, recombine with other pieces of endogenous DNA through shared homologous regions. If this complementary strand is attached to an oligonucleotide that contains a mutation or a different sequence or an additional nucleotide, it too is incorporated into the newly synthesized strand as a result of the recombination. As a result of the proofreading function, it is possible for the new sequence of DNA to serve as the template. Thus, the transferred DNA is incorporated into the genome.

Attached to these pieces of targeting DNA are regions of DNA that may interact with or control the expression of a HER-2sv polypeptide, *e.g.*, flanking sequences. For example, a promoter/enhancer element, a suppressor, or an exogenous transcription modulatory element is inserted in the genome of the intended host cell in proximity and orientation sufficient to influence the transcription of DNA encoding the desired HER-2sv polypeptide. The control element controls a portion of the DNA present in the host cell genome. Thus, the expression of the desired HER-2sv polypeptide may be achieved not by transfection of DNA that encodes the HER-2sv gene itself, but rather by the use of targeting DNA (containing regions of homology with the endogenous gene of interest) coupled with DNA regulatory segments that provide the endogenous gene sequence with recognizable signals for transcription of a HER-2sv gene.

In an exemplary method, the expression of a desired targeted gene in a cell (*i.e.*, a desired endogenous cellular gene) is altered via homologous recombination into the cellular genome at a preselected site, by the introduction of DNA that includes at least a regulatory sequence, an exon, and a splice donor site. These components are introduced into the chromosomal (genomic) DNA in such a manner that this, in effect, results in the production of a new transcription unit (in which the regulatory sequence, the exon, and the splice donor site present in the DNA construct

are operatively linked to the endogenous gene). As a result of the introduction of these components into the chromosomal DNA, the expression of the desired endogenous gene is altered.

Altered gene expression, as described herein, encompasses activating (or causing to be expressed) a gene which is normally silent (unexpressed) in the cell as obtained, as well as increasing the expression of a gene which is not expressed at physiologically significant levels in the cell as obtained. The embodiments further encompass changing the pattern of regulation or induction such that it is different from the pattern of regulation or induction that occurs in the cell as obtained, and reducing (including eliminating) the expression of a gene which is expressed in the cell as obtained.

One method by which homologous recombination can be used to increase, or cause, HER-2sv polypeptide production from a cell's endogenous HER-2sv gene involves first using homologous recombination to place a recombination sequence from a site-specific recombination system (*e.g.*, Cre/loxP, FLP/FRT) (Sauer, 1994, *Curr. Opin. Biotechnol.*, 5:521-27; Sauer, 1993, *Methods Enzymol.*, 225:890-900) upstream of (*i.e.*, 5' to) the cell's endogenous genomic HER-2sv polypeptide coding region. A plasmid containing a recombination site homologous to the site that was placed just upstream of the genomic HER-2sv polypeptide coding region is introduced into the modified cell line along with the appropriate recombinase enzyme. This recombinase causes the plasmid to integrate, via the plasmid's recombination site, into the recombination site located just upstream of the genomic HER-2sv polypeptide coding region in the cell line (Baubonis and Sauer, 1993, *Nucleic Acids Res.* 21:2025-29; O'Gorman *et al.*, 1991, *Science* 251:1351-55). Any flanking sequences known to increase transcription (*e.g.*, enhancer/promoter, intron, translational enhancer), if properly positioned in this plasmid, would integrate in such a manner as to create a new or modified transcriptional unit resulting in *de novo* or increased HER-2sv polypeptide production from the cell's endogenous HER-2sv gene.

A further method to use the cell line in which the site specific recombination sequence had been placed just upstream of the cell's endogenous genomic HER-2sv polypeptide coding region is to use homologous recombination to introduce a second recombination site elsewhere in the cell line's genome. The appropriate recombinase enzyme is then introduced into the two-recombination-site cell line, causing a

recombination event (deletion, inversion, and translocation) (Sauer, 1994, *Curr. Opin. Biotechnol.*, 5:521-27; Sauer, 1993, *Methods Enzymol.*, 225:890-900) that would create a new or modified transcriptional unit resulting in *de novo* or increased HER-2sv polypeptide production from the cell's endogenous HER-2sv gene.

An additional approach for increasing, or causing, the expression of HER-2sv polypeptide from a cell's endogenous HER-2sv gene involves increasing, or causing, the expression of a gene or genes (*e.g.*, transcription factors) and/or decreasing the expression of a gene or genes (*e.g.*, transcriptional repressors) in a manner which results in *de novo* or increased HER-2sv polypeptide production from the cell's endogenous HER-2sv gene. This method includes the introduction of a non-naturally occurring polypeptide (*e.g.*, a polypeptide comprising a site specific DNA binding domain fused to a transcriptional factor domain) into the cell such that *de novo* or increased HER-2sv polypeptide production from the cell's endogenous HER-2sv gene results.

The present invention further relates to DNA constructs useful in the method of altering expression of a target gene. In certain embodiments, the exemplary DNA constructs comprise: (a) one or more targeting sequences, (b) a regulatory sequence, (c) an exon, and (d) an unpaired splice-donor site. The targeting sequence in the DNA construct directs the integration of elements (a) - (d) into a target gene in a cell such that the elements (b) - (d) are operatively linked to sequences of the endogenous target gene. In another embodiment, the DNA constructs comprise: (a) one or more targeting sequences, (b) a regulatory sequence, (c) an exon, (d) a splice-donor site, (e) an intron, and (f) a splice-acceptor site, wherein the targeting sequence directs the integration of elements (a) - (f) such that the elements of (b) - (f) are operatively linked to the endogenous gene. The targeting sequence is homologous to the preselected site in the cellular chromosomal DNA with which homologous recombination is to occur. In the construct, the exon is generally 3' of the regulatory sequence and the splice-donor site is 3' of the exon.

If the sequence of a particular gene is known, such as the nucleic acid sequence of HER-2sv polypeptide presented herein, a piece of DNA that is complementary to a selected region of the gene can be synthesized or otherwise obtained, such as by appropriate restriction of the native DNA at specific recognition sites bounding the region of interest. This piece serves as a targeting sequence upon insertion into the cell and will hybridize to its homologous region within the genome.

If this hybridization occurs during DNA replication, this piece of DNA, and any additional sequence attached thereto, will act as an Okazaki fragment and will be incorporated into the newly synthesized daughter strand of DNA. The present invention, therefore, includes nucleotides encoding a HER-2sv polypeptide, which nucleotides may be used as targeting sequences.

HER-2sv polypeptide cell therapy, *e.g.*, the implantation of cells producing HER-2sv polypeptides, is also contemplated. This embodiment involves implanting cells capable of synthesizing and secreting a biologically active form of HER-2sv polypeptide. Such HER-2sv polypeptide-producing cells can be cells that are natural producers of HER-2sv polypeptides or may be recombinant cells whose ability to produce HER-2sv polypeptides has been augmented by transformation with a gene encoding the desired HER-2sv polypeptide or with a gene augmenting the expression of HER-2sv polypeptide. Such a modification may be accomplished by means of a vector suitable for delivering the gene as well as promoting its expression and secretion. In order to minimize a potential immunological reaction in patients being administered a HER-2sv polypeptide, as may occur with the administration of a polypeptide of a foreign species, it is preferred that the natural cells producing HER-2sv polypeptide be of human origin and produce human HER-2sv polypeptide. Likewise, it is preferred that the recombinant cells producing HER-2sv polypeptide be transformed with an expression vector containing a gene encoding a human HER-2sv polypeptide.

Implanted cells may be encapsulated to avoid the infiltration of surrounding tissue. Human or non-human animal cells may be implanted in patients in biocompatible, semipermeable polymeric enclosures or membranes that allow the release of HER-2sv polypeptide, but that prevent the destruction of the cells by the patient's immune system or by other detrimental factors from the surrounding tissue. Alternatively, the patient's own cells, transformed to produce HER-2sv polypeptides *ex vivo*, may be implanted directly into the patient without such encapsulation.

Techniques for the encapsulation of living cells are known in the art, and the preparation of the encapsulated cells and their implantation in patients may be routinely accomplished. For example, Baetge *et al.* (International Pub. No. WO 95/05452 and International App. No. PCT/US94/09299) describe membrane capsules containing genetically engineered cells for the effective delivery of biologically active molecules. The capsules are biocompatible and are easily retrievable. The capsules

encapsulate cells transfected with recombinant DNA molecules comprising DNA sequences coding for biologically active molecules operatively linked to promoters that are not subject to down-regulation *in vivo* upon implantation into a mammalian host. The devices provide for the delivery of the molecules from living cells to specific sites within a recipient. In addition, *see* U.S. Patent Nos. 4,892,538; 5,011,472; and 5,106,627. A system for encapsulating living cells is described in International Pub. No. WO 91/10425 (Aebischer *et al.*). *See also*, International Pub. No. WO 91/10470 (Aebischer *et al.*); Winn *et al.*, 1991, *Exper. Neurol.* 113:322-29; Aebischer *et al.*, 1991, *Exper. Neurol.* 111:269-75; and Tresco *et al.*, 1992, *ASAIO* 38:17-23.

In vivo and *in vitro* gene therapy delivery of HER-2sv polypeptides is also envisioned. One example of a gene therapy technique is to use the HER-2sv gene (either genomic DNA, cDNA, and/or synthetic DNA) encoding a HER-2sv polypeptide that may be operably linked to a constitutive or inducible promoter to form a "gene therapy DNA construct." The promoter may be homologous or heterologous to the endogenous HER-2sv gene, provided that it is active in the cell or tissue type into which the construct will be inserted. Other components of the gene therapy DNA construct may optionally include DNA molecules designed for site-specific integration (*e.g.*, endogenous sequences useful for homologous recombination), tissue-specific promoters, enhancers or silencers, DNA molecules capable of providing a selective advantage over the parent cell, DNA molecules useful as labels to identify transformed cells, negative selection systems, cell specific binding agents (as, for example, for cell targeting), cell-specific internalization factors, transcription factors enhancing expression from a vector, and factors enabling vector production.

A gene therapy DNA construct can then be introduced into cells (either *ex vivo* or *in vivo*) using viral or non-viral vectors. One means for introducing the gene therapy DNA construct is by means of viral vectors as described herein. Certain vectors, such as retroviral vectors, will deliver the DNA construct to the chromosomal DNA of the cells, and the gene can integrate into the chromosomal DNA. Other vectors will function as episomes, and the gene therapy DNA construct will remain in the cytoplasm.

In yet other embodiments, regulatory elements can be included for the controlled expression of the HER-2sv gene in the target cell. Such elements are

turned on in response to an appropriate effector. In this way, a therapeutic polypeptide can be expressed when desired. One conventional control means involves the use of small molecule dimerizers or rapalogs to dimerize chimeric proteins which contain a small molecule-binding domain and a domain capable of initiating a biological process, such as a DNA-binding protein or transcriptional activation protein (*see* International Pub. Nos. WO 96/41865, WO 97/31898, and WO 97/31899). The dimerization of the proteins can be used to initiate transcription of the transgene.

An alternative regulation technology uses a method of storing proteins expressed from the gene of interest inside the cell as an aggregate or cluster. The gene of interest is expressed as a fusion protein that includes a conditional aggregation domain that results in the retention of the aggregated protein in the endoplasmic reticulum. The stored proteins are stable and inactive inside the cell. The proteins can be released, however, by administering a drug (*e.g.*, small molecule ligand) that removes the conditional aggregation domain and thereby specifically breaks apart the aggregates or clusters so that the proteins may be secreted from the cell. *See* Aridor *et al.*, 2000, *Science* 287:816-17 and Rivera *et al.*, 2000, *Science* 287:826-30.

Other suitable control means or gene switches include, but are not limited to, the systems described herein. Mifepristone (RU486) is used as a progesterone antagonist. The binding of a modified progesterone receptor ligand-binding domain to the progesterone antagonist activates transcription by forming a dimer of two transcription factors that then pass into the nucleus to bind DNA. The ligand-binding domain is modified to eliminate the ability of the receptor to bind to the natural ligand. The modified steroid hormone receptor system is further described in U.S. Patent No. 5,364,791 and International Pub. Nos. WO 96/40911 and WO 97/10337.

Yet another control system uses ecdysone (a fruit fly steroid hormone), which binds to and activates an ecdysone receptor (cytoplasmic receptor). The receptor then translocates to the nucleus to bind a specific DNA response element (promoter from ecdysone-responsive gene). The ecdysone receptor includes a transactivation domain, DNA-binding domain, and ligand-binding domain to initiate transcription. The ecdysone system is further described in U.S. Patent No. 5,514,578 and International Pub. Nos. WO 97/38117, WO 96/37609, and WO 93/03162.

Another control means uses a positive tetracycline-controllable transactivator. This system involves a mutated tet repressor protein DNA-binding domain (mutated tet R-4 amino acid changes which resulted in a reverse tetracycline-regulated transactivator protein, *i.e.*, it binds to a tet operator in the presence of tetracycline) linked to a polypeptide which activates transcription. Such systems are described in U.S. Patent Nos. 5,464,758, 5,650,298, and 5,654,168.

Additional expression control systems and nucleic acid constructs are described in U.S. Patent Nos. 5,741,679 and 5,834,186, to Innovir Laboratories Inc.

In vivo gene therapy may be accomplished by introducing the gene encoding HER-2sv polypeptide into cells via local injection of a HER-2sv nucleic acid molecule or by other appropriate viral or non-viral delivery vectors. Hefti 1994, *Neurobiology* 25:1418-35. For example, a nucleic acid molecule encoding a HER-2sv polypeptide may be contained in an adeno-associated virus (AAV) vector for delivery to the targeted cells (*see, e.g.*, Johnson, International Pub. No. WO 95/34670; International App. No. PCT/US95/07178). The recombinant AAV genome typically contains AAV inverted terminal repeats flanking a DNA sequence encoding a HER-2sv polypeptide operably linked to functional promoter and polyadenylation sequences.

Alternative suitable viral vectors include, but are not limited to, retrovirus, adenovirus, herpes simplex virus, lentivirus, hepatitis virus, parvovirus, papovavirus, poxvirus, alphavirus, coronavirus, rhabdovirus, paramyxovirus, and papilloma virus vectors. U.S. Patent No. 5,672,344 describes an *in vivo* viral-mediated gene transfer system involving a recombinant neurotrophic HSV-1 vector. U.S. Patent No. 5,399,346 provides examples of a process for providing a patient with a therapeutic protein by the delivery of human cells that have been treated *in vitro* to insert a DNA segment encoding a therapeutic protein. Additional methods and materials for the practice of gene therapy techniques are described in U.S. Patent Nos. 5,631,236 (involving adenoviral vectors), 5,672,510 (involving retroviral vectors), 5,635,399 (involving retroviral vectors expressing cytokines).

Nonviral delivery methods include, but are not limited to, liposome-mediated transfer, naked DNA delivery (direct injection), receptor-mediated transfer (ligand-DNA complex), electroporation, calcium phosphate precipitation, and microparticle bombardment (*e.g.*, gene gun). Gene therapy materials and methods may also include inducible promoters, tissue-specific enhancer-promoters, DNA sequences designed

for site-specific integration, DNA sequences capable of providing a selective advantage over the parent cell, labels to identify transformed cells, negative selection systems and expression control systems (safety measures), cell-specific binding agents (for cell targeting), cell-specific internalization factors, and transcription factors to enhance expression by a vector as well as methods of vector manufacture. Such additional methods and materials for the practice of gene therapy techniques are described in U.S. Patent Nos. 4,970,154 (involving electroporation techniques), 5,679,559 (describing a lipoprotein-containing system for gene delivery), 5,676,954 (involving liposome carriers), 5,593,875 (describing methods for calcium phosphate transfection), and 4,945,050 (describing a process wherein biologically active particles are propelled at cells at a speed whereby the particles penetrate the surface of the cells and become incorporated into the interior of the cells), and International Pub. No. WO 96/40958 (involving nuclear ligands).

It is also contemplated that HER-2sv gene therapy or cell therapy can further include the delivery of one or more additional polypeptide(s) in the same or a different cell(s). Such cells may be separately introduced into the patient, or the cells may be contained in a single implantable device, such as the encapsulating membrane described above, or the cells may be separately modified by means of viral vectors.

A means to increase endogenous HER-2sv polypeptide expression in a cell via gene therapy is to insert one or more enhancer elements into the HER-2sv polypeptide promoter, where the enhancer elements can serve to increase transcriptional activity of the HER-2sv gene. The enhancer elements used will be selected based on the tissue in which one desires to activate the gene – enhancer elements known to confer promoter activation in that tissue will be selected. For example, if a gene encoding a HER-2sv polypeptide is to be "turned on" in T-cells, the *lck* promoter enhancer element may be used. Here, the functional portion of the transcriptional element to be added may be inserted into a fragment of DNA containing the HER-2sv polypeptide promoter (and optionally, inserted into a vector and/or 5' and/or 3' flanking sequences) using standard cloning techniques. This construct, known as a "homologous recombination construct," can then be introduced into the desired cells either *ex vivo* or *in vivo*.

Gene therapy also can be used to decrease HER-2sv polypeptide expression by modifying the nucleotide sequence of the endogenous promoter. Such modification is typically accomplished via homologous recombination methods. For

example, a DNA molecule containing all or a portion of the promoter of the HER-2sv gene selected for inactivation can be engineered to remove and/or replace pieces of the promoter that regulate transcription. For example, the TATA box and/or the binding site of a transcriptional activator of the promoter may be deleted using standard molecular biology techniques; such deletion can inhibit promoter activity thereby repressing the transcription of the corresponding HER-2sv gene. The deletion of the TATA box or the transcription activator binding site in the promoter may be accomplished by generating a DNA construct comprising all or the relevant portion of the HER-2sv polypeptide promoter (from the same or a related species as the HER-2sv gene to be regulated) in which one or more of the TATA box and/or transcriptional activator binding site nucleotides are mutated via substitution, deletion and/or insertion of one or more nucleotides. As a result, the TATA box and/or activator binding site has decreased activity or is rendered completely inactive. This construct, which also will typically contain at least about 500 bases of DNA that correspond to the native (endogenous) 5' and 3' DNA sequences adjacent to the promoter segment that has been modified, may be introduced into the appropriate cells (either *ex vivo* or *in vivo*) either directly or via a viral vector as described herein. Typically, the integration of the construct into the genomic DNA of the cells will be via homologous recombination, where the 5' and 3' DNA sequences in the promoter construct can serve to help integrate the modified promoter region via hybridization to the endogenous chromosomal DNA.

15. Therapeutic Uses

HER-2sv nucleic acid molecules, polypeptides, and antagonists thereof can be used to treat, diagnose, ameliorate, or prevent a number of diseases, disorders, or conditions, including those recited herein.

HER-2sv polypeptide antagonists include those molecules that regulate HER-2sv polypeptide activity by decreasing at least one activity of the mature form of the HER-2sv polypeptide. Antagonists may be co-factors, such as a protein, peptide, carbohydrate, lipid, or small molecular weight molecule, which interact with HER-2sv polypeptide and thereby regulate its activity. Potential polypeptide antagonists include antibodies that react with either soluble or membrane-bound forms of HER-2sv polypeptides that comprise part or all of the extracellular domains of the said proteins. Molecules that regulate HER-2sv polypeptide expression typically include

nucleic acids encoding HER-2sv polypeptide that can act as anti-sense regulators of expression.

Since aberrant HER-2 expression has been detected in a number of human cancers, including breast, ovarian, gastric, lung, and oral cancer, HER-2sv nucleic acid molecules, polypeptides, and antagonists thereof may be useful in diagnosing or treating diseases including breast, ovarian, gastric, lung, and oral cancer. Other human carcinomas involving HER-2sv polypeptides are encompassed within the scope of this invention.

Antagonists of HER-2sv polypeptide function may be used (simultaneously or sequentially) in combination with one or more cytokines, growth factors, antibiotics, anti-inflammatories, or chemotherapeutic agents as are appropriate for the condition being treated.

Other diseases or disorders caused by or mediated by undesirable levels of HER-2sv polypeptides are encompassed within the scope of the invention. Undesirable levels preferably include excessive levels of HER-2sv polypeptides.

16. Uses of HER-2sv Nucleic Acids and Polypeptides

Nucleic acid molecules of the invention (including those that do not themselves encode biologically active polypeptides) may be used to map the locations of the HER-2sv gene and related genes on chromosomes. Mapping may be done by techniques known in the art, such as PCR amplification and *in situ* hybridization.

HER-2sv nucleic acid molecules (including those that do not themselves encode biologically active polypeptides), may be useful as hybridization probes in diagnostic assays to test, either qualitatively or quantitatively, for the presence of a HER-2sv nucleic acid molecule in mammalian tissue or bodily fluid samples.

Other methods may also be employed where it is desirable to inhibit the activity of one or more HER-2sv polypeptides. Such inhibition may be effected by nucleic acid molecules that are complementary to and hybridize to expression control sequences (triple helix formation) or to HER-2sv mRNA. For example, antisense DNA or RNA molecules, which have a sequence that is complementary to at least a portion of a HER-2sv gene can be introduced into the cell. Anti-sense probes may be designed by available techniques using the sequence of the HER-2sv gene disclosed herein. Typically, each such antisense molecule will be complementary to the start site (5' end) of each selected HER-2sv gene. When the antisense molecule then

hybridizes to the corresponding HER-2sv mRNA, translation of this mRNA is prevented or reduced. Anti-sense inhibitors provide information relating to the decrease or absence of a HER-2sv polypeptide in a cell or organism.

Alternatively, gene therapy may be employed to create a dominant-negative inhibitor of one or more HER-2sv polypeptides. In this situation, the DNA encoding a mutant polypeptide of each selected HER-2sv polypeptide can be prepared and introduced into the cells of a patient using either viral or non-viral methods as described herein. Each such mutant is typically designed to compete with endogenous polypeptide in its biological role.

In addition, a HER-2sv polypeptide, whether biologically active or not, may be used as an immunogen, that is, the polypeptide contains at least one epitope to which antibodies may be raised. Selective binding agents that bind to a HER-2sv polypeptide (as described herein) may be used for *in vivo* and *in vitro* diagnostic purposes, including, but not limited to, use in labeled form to detect the presence of HER-2sv polypeptide in a body fluid or cell sample. The antibodies may also be used to prevent, treat, or diagnose a number of diseases and disorders, including those recited herein. The antibodies may bind to a HER-2sv polypeptide so as to diminish or block at least one activity characteristic of a HER-2sv polypeptide, or may bind to a polypeptide to increase at least one activity characteristic of a HER-2sv polypeptide (including by increasing the pharmacokinetics of the HER-2sv polypeptide).

The human HER-2sv nucleic acids of the present invention are also useful tools for isolating the corresponding chromosomal HER-2sv polypeptide genes. The human HER-2sv genomic DNA can be used to identify heritable tissue-degenerating diseases.

The following examples are intended for illustration purposes only, and should not be construed as limiting the scope of the invention in any way.

EXAMPLE 1

Cloning of HER-2 Splice Variants

Generally, materials and methods as described in Sambrook *et al. supra* were used to clone and analyze genes encoding rat HER-2sv polypeptides.

To isolate HER-2 splice variant cDNA sequences, a proprietary human tissue cDNA library array was screened by PCR using the amplimers 2771-31 (5'-C-G-G-T-C-G-A-C-G-A-G-C-T-C-G-A-G-G-G-T-C-3'; SEQ ID NO: 14) and 2771-33 (5'-C-

A-G-T-C-T-C-C-G-C-A-T-C-G-T-G-T-A-C-T-T-C-C-G-3'; SEQ ID NO: 15). PCR amplications were prepared using either 100 ng of human tissue cDNA template, 10 ng of Clontech Marathon human cDNA template, or 10 ng of Clontech Marathon human xenograft cDNA template; 10 pmol of amplimers; the Advantage-HF2 PCR kit (Clontech); and 2 µl of GC-Melt (Clontech) in final volume a 50 µl final. Reactions were performed at 94°C for 2 minutes for one cycle; 94°C for 30 seconds, 65°C for 30 seconds, and 72°C for 3 minutes for 40 cycles; and 72°C for 7 minutes for one cycle.

The products generated in this first PCR were reamplified in nested PCR amplifications using 1 ml of the product from the first PCR and the amplimers 2771-32 (5'-G-A-G-C-C-G-C-A-G-T-G-A-G-C-A-C-C-A-T-G-G-A-G-3'; SEQ ID NO: 16) and 2771-34 (5'-G-C-T-G-C-C-G-T-C-G-C-T-T-G-A-T-G-A-G-G-A-T-C-3'; SEQ ID NO: 17) and the same conditions employed in the initial PCR. PCR products generated in the nested PCR amplifications were analyzed by gel elctrophoresis, with products of various sizes being detected in the following cDNA libraries: fetal stomach, fetal pancreas, fetal kidney, fetal lung, fetal heart, uterus, testis, placenta, fetal scalp, fetal calveria, spinal column, trachea, lung tumor, T1543 breast tumor, ovary tumor, colon tumor, prostate tumor, fetal small, intestine, and mononuclear circulating lymphocytes. These products were ligated into the vector pGEM-T EASY and used to transform *E .coli*. Plasmid DNA was isolated from six individual colonies isolated from each transformation and the inserts were sequenced.

Five splice variants of the extracellular domain of the HER-2 receptor tyrosine kinase gene were identified in the PCR screens. Figures 6A-6C illustrate the amino acid sequence alignment of the extracellular portion of human HER-2 (SEQ ID NO: 11) and the polypeptides encoded by the five splice variants. These splice variants included HER-2sv form 97 (SEQ ID NO: 4), HER-2sv 184 (SEQ ID NO: 10), HER-2sv 119 (SEQ ID NO: 6), HER-2sv 68 (SEQ ID NO: 2), and HER-2sv 156 (SEQ ID NO: 8).

Figure 7 illustrates a schematic representation of the structure of the known form of the extracellular domain of the HER-2 gene and human HER-2sv forms 119, 184, 97, 68, and 156. The known form of the HER-2 extracellular domain consists of 17 exons. Structurally, it possesses two receptor L-domains, a furin-like domain, and a transmembrane domain. The receptor L-domains are ligand-binding domains, each such domain consisting of a single-stranded right hand beta-helix. The furin-like

domain is a cysteine-rich region, which is found in a variety of proteins that are involved in signal transduction.

In HER-2sv form 119, an additional exon between exons 9 and 10 of the known form of the HER-2 extracellular domain encodes an additional 39 amino acids. This additional sequence disrupts the second L-domain of HER-2. Form 119 sequences were detected in the scalp cDNA library.

In HER-2sv form 184, an additional exon between exons 14 and 15 of the known form of the HER-2 extracellular domain encodes an additional 34 amino acids. This additional sequence does not disrupt the L-domains or furin-like domain of HER-2. Form 184 sequences were detected in the trachea cDNA library.

HER-2sv form 97 lacks exon 16 of the known form of the HER-2 extracellular domain. The deletion of exon 16 does not disrupt the L-domains or furin-like domain of HER-2. Form 97 sequences were detected in both the calveria and trachea cDNA libraries.

HER-2sv form 68 possesses atypical splice sites within exons 7 and 12 of the known form of the HER-2 extracellular domain, resulting in a 187 amino acid deletion. The deletion of this portion of HER-2 removes most of the furin-like domain and most of the second receptor L-domain. Form 68 sequences were detected in the fetal kidney, testis, colon tumor, and T1543 breast carcinoma cDNA libraries.

HER-2sv form 156 possesses atypical splice sites within exons 4 and 15 of the known form of the HER-2 extracellular domain. The atypical splice site in exon 4 generates a frame shift and premature stop codon. This splice variant lacks both the furin-like domain and second receptor L-domain. Form 156 sequences were detected in the T1543 breast carcinoma cDNA library.

EXAMPLE 2

HER-2sv mRNA Expression

The expression of HER-2sv mRNA is examined by Northern blot analysis. Multiple human tissue northern blots (Clontech) are probed with a suitable restriction fragment isolated from a human HER-2sv polypeptide cDNA clone. The probe is labeled with ^{32}P-dCTP using standard techniques.

Northern blots are prehybridized for 2 hours at 42°C in hybridization solution (5X SSC, 50% deionized formamide, 5X Denhardt's solution, 0.5% SDS, and 100 mg/ml denatured salmon sperm DNA) and then hybridized at 42°C overnight in fresh

hybridization solution containing 5 ng/ml of the labeled probe. Following hybridization, the filters are washed twice for 10 minutes at room temperature in 2X SSC and 0.1% SDS, and then twice for 30 minutes at 65°C in 0.1X SSC and 0.1% SDS. The blots are then exposed to autoradiography.

The expression of HER-2sv mRNA is localized by *in situ* hybridization. A panel of normal embryonic and adult mouse tissues is fixed in 4% paraformaldehyde, embedded in paraffin, and sectioned at 5 μm. Sectioned tissues are permeabilized in 0.2 M HCl, digested with Proteinase K, and acetylated with triethanolamine and acetic anhydride. Sections are prehybridized for 1 hour at 60°C in hybridization solution (300 mM NaCl, 20 mM Tris-HCl, pH 8.0, 5 mM EDTA, 1X Denhardt's solution, 0.2% SDS, 10 mM DTT, 0.25 mg/ml tRNA, 25 μg/ml polyA, 25 μg/ml polyC and 50% formamide) and then hybridized overnight at 60°C in the same solution containing 10% dextran and 2 x 10^4 cpm/μl of a ^{33}P-labeled antisense riboprobe complementary to the human HER-2sv gene. The riboprobe is obtained by *in vitro* transcription of a clone containing human HER-2sv cDNA sequences using standard techniques.

Following hybridization, sections are rinsed in hybridization solution, treated with RNaseA to digest unhybridized probe, and then washed in 0.1X SSC at 55°C for 30 minutes. Sections are then immersed in NTB-2 emulsion (Kodak, Rochester, NY), exposed for 3 weeks at 4°C, developed, and counterstained with hematoxylin and eosin. Tissue morphology and hybridization signal are simultaneously analyzed by darkfield and standard illumination for brain (one sagittal and two coronal sections), gastrointestinal tract (esophagus, stomach, duodenum, jejunum, ileum, proximal colon, and distal colon), pituitary, liver, lung, heart, spleen, thymus, lymph nodes, kidney, adrenal, bladder, pancreas, salivary gland, male and female reproductive organs (ovary, oviduct, and uterus in the female; and testis, epididymus, prostate, seminal vesicle, and vas deferens in the male), BAT and WAT (subcutaneous, peri-renal), bone (femur), skin, breast, and skeletal muscle.

EXAMPLE 3

Production of HER-2sv Polypeptides

A. Expression of HER-2sv Polypeptides in Bacteria

PCR is used to amplify template DNA sequences encoding a HER-2sv polypeptide using primers corresponding to the 5' and 3' ends of the sequence. The amplified DNA products may be modified to contain restriction enzyme sites to allow for insertion into expression vectors. PCR products are gel purified and inserted into expression vectors using standard recombinant DNA methodology. An exemplary vector, such as pAMG21 (ATCC no. 98113) containing the lux promoter and a gene encoding kanamycin resistance is digested with Bam HI and Nde I for directional cloning of inserted DNA. The ligated mixture is transformed into an *E. coli* host strain by electroporation and transformants are selected for kanamycin resistance. Plasmid DNA from selected colonies is isolated and subjected to DNA sequencing to confirm the presence of the insert.

Transformed host cells are incubated in 2xYT medium containing 30 μg/mL kanamycin at 30°C prior to induction. Gene expression is induced by the addition of N-(3-oxohexanoyl)-dl-homoserine lactone to a final concentration of 30 ng/mL followed by incubation at either 30°C or 37°C for six hours. The expression of HER-2sv polypeptide is evaluated by centrifugation of the culture, resuspension and lysis of the bacterial pellets, and analysis of host cell proteins by SDS-polyacrylamide gel electrophoresis.

Inclusion bodies containing HER-2sv polypeptide are purified as follows. Bacterial cells are pelleted by centrifugation and resuspended in water. The cell suspension is lysed by sonication and pelleted by centrifugation at 195,000 x*g* for 5 to 10 minutes. The supernatant is discarded, and the pellet is washed and transferred to a homogenizer. The pellet is homogenized in 5 mL of a Percoll solution (75% liquid Percoll and 0.15 M NaCl) until uniformly suspended and then diluted and centrifuged at 21,600 x*g* for 30 minutes. Gradient fractions containing the inclusion bodies are recovered and pooled. The isolated inclusion bodies are analyzed by SDS-PAGE.

A single band on an SDS polyacrylamide gel corresponding to *E. coli*-produced HER-2sv polypeptide is excised from the gel, and the N-terminal amino acid sequence is determined essentially as described by Matsudaira *et al.*, 1987, *J. Biol. Chem.* 262:10-35.

B. Expression of HER-2sv Polypeptide in Mammalian Cells

PCR is used to amplify template DNA sequences encoding a HER-2sv polypeptide using primers corresponding to the 5' and 3' ends of the sequence. The

amplified DNA products may be modified to contain restriction enzyme sites to allow for insertion into expression vectors. PCR products are gel purified and inserted into expression vectors using standard recombinant DNA methodology. An exemplary expression vector, pCEP4 (Invitrogen, Carlsbad, CA), that contains an Epstein-Barr virus origin of replication, may be used for the expression of HER-2sv polypeptides in 293-EBNA-1 cells. Amplified and gel purified PCR products are ligated into pCEP4 vector and introduced into 293-EBNA cells by lipofection. The transfected cells are selected in 100 μg/mL hygromycin and the resulting drug-resistant cultures are grown to confluence. The cells are then cultured in serum-free media for 72 hours. The conditioned media is removed and HER-2sv polypeptide expression is analyzed by SDS-PAGE.

HER-2sv polypeptide expression may be detected by silver staining. Alternatively, HER-2sv polypeptide is produced as a fusion protein with an epitope tag, such as an IgG constant domain or a FLAG epitope, which may be detected by Western blot analysis using antibodies to the peptide tag.

HER-2sv polypeptides may be excised from an SDS-polyacrylamide gel, or HER-2sv fusion proteins are purified by affinity chromatography to the epitope tag, and subjected to N-terminal amino acid sequence analysis as described herein.

C. Expression and Purification of HER-2sv Polypeptide in Mammalian Cells

HER-2sv polypeptide expression constructs are introduced into 293 EBNA or CHO cells using either a lipofection or calcium phosphate protocol.

To conduct functional studies on the HER-2sv polypeptides that are produced, large quantities of conditioned media are generated from a pool of hygromycin selected 293 EBNA clones. The cells are cultured in 500 cm Nunc Triple Flasks to 80% confluence before switching to serum free media a week prior to harvesting the media. Conditioned media is harvested and frozen at -20°C until purification.

Conditioned media is purified by affinity chromatography as described below. The media is thawed and then passed through a 0.2 μm filter. A Protein G column is equilibrated with PBS at pH 7.0, and then loaded with the filtered media. The column is washed with PBS until the absorbance at A_{280} reaches a baseline. HER-2sv polypeptide is eluted from the column with 0.1 M Glycine-HCl at pH 2.7 and immediately neutralized with 1 M Tris-HCl at pH 8.5. Fractions containing HER-2sv polypeptide are pooled, dialyzed in PBS, and stored at -70°C.

For Factor Xa cleavage of the human HER-2sv polypeptide-Fc fusion polypeptide, affinity chromatography-purified protein is dialyzed in 50 mM Tris-HCl, 100 mM NaCl, 2 mM $CaCl_2$ at pH 8.0. The restriction protease Factor Xa is added to the dialyzed protein at 1/100 (w/w) and the sample digested overnight at room temperature.

EXAMPLE 4

Production of Anti-HER-2sv Polypeptide Antibodies

Antibodies to HER-2sv polypeptides may be obtained by immunization with purified protein or with HER-2sv peptides produced by biological or chemical synthesis. Suitable procedures for generating antibodies include those described in Hudson and Bay, *Practical Immunology* (2nd ed., Blackwell Scientific Publications).

In one procedure for the production of antibodies, animals (typically mice or rabbits) are injected with a HER-2sv antigen (such as a HER-2sv polypeptide), and those with sufficient serum titer levels as determined by ELISA are selected for hybridoma production. Spleens of immunized animals are collected and prepared as single cell suspensions from which splenocytes are recovered. The splenocytes are fused to mouse myeloma cells (such as Sp2/0-Ag14 cells), are first incubated in DMEM with 200 U/mL penicillin, 200 μg/mL streptomycin sulfate, and 4 mM glutamine, and are then incubated in HAT selection medium (hypoxanthine, aminopterin, and thymidine). After selection, the tissue culture supernatants are taken from each fusion well and tested for anti-HER-2sv antibody production by ELISA.

Alternative procedures for obtaining anti-HER-2sv antibodies may also be employed, such as the immunization of transgenic mice harboring human Ig loci for production of human antibodies, and the screening of synthetic antibody libraries, such as those generated by mutagenesis of an antibody variable domain.

EXAMPLE 5

Expression of HER-2sv Polypeptide in Transgenic Mice

To assess the biological activity of HER-2sv polypeptide, a construct encoding a HER-2sv polypeptide/Fc fusion protein under the control of a liver specific ApoE promoter is prepared. The delivery of this construct is expected to cause pathological changes that are informative as to the function of HER-2sv polypeptide. Similarly, a construct containing the full-length HER-2sv polypeptide under the control of the beta

actin promoter is prepared. The delivery of this construct is expected to result in ubiquitous expression.

To generate these constructs, PCR is used to amplify template DNA sequences encoding a HER-2sv polypeptide using primers that correspond to the 5' and 3' ends of the desired sequence and which incorporate restriction enzyme sites to permit insertion of the amplified product into an expression vector. Following amplification, PCR products are gel purified, digested with the appropriate restriction enzymes, and ligated into an expression vector using standard recombinant DNA techniques. For example, amplified HER-2sv polypeptide sequences can be cloned into an expression vector under the control of the human β-actin promoter as described by Graham *et al.*, 1997, *Nature Genetics*, 17:272-74 and Ray *et al.*, 1991, *Genes Dev.* 5:2265-73.

Following ligation, reaction mixtures are used to transform an *E. coli* host strain by electroporation and transformants are selected for drug resistance. Plasmid DNA from selected colonies is isolated and subjected to DNA sequencing to confirm the presence of an appropriate insert and absence of mutation. The HER-2sv polypeptide expression vector is purified through two rounds of CsCl density gradient centrifugation, cleaved with a suitable restriction enzyme, and the linearized fragment containing the HER-2sv polypeptide transgene is purified by gel electrophoresis. The purified fragment is resuspended in 5 mM Tris, pH 7.4, and 0.2 mM EDTA at a concentration of 2 mg/mL.

Single-cell embryos from BDF1 x BDF1 bred mice are injected as described (International Pub. No. WO 97/23614). Embryos are cultured overnight in a CO_2 incubator and 15-20 two-cell embryos are transferred to the oviducts of a pseudopregnant CD1 female mice. Offspring obtained from the implantation of microinjected embryos are screened by PCR amplification of the integrated transgene in genomic DNA samples as follows. Ear pieces are digested in 20 mL ear buffer (20 mM Tris, pH 8.0, 10 mM EDTA, 0.5% SDS, and 500 mg/mL proteinase K) at 55°C overnight. The sample is then diluted with 200 mL of TE, and 2 mL of the ear sample is used in a PCR reaction using appropriate primers.

At 8 weeks of age, transgenic founder animals and control animals are sacrificed for necropsy and pathological analysis. Portions of spleen are removed and total cellular RNA isolated from the spleens using the Total RNA Extraction Kit (Qiagen) and transgene expression determined by RT-PCR. RNA recovered from spleens is converted to cDNA using the SuperScript™ Preamplification System

(Gibco-BRL) as follows. A suitable primer, located in the expression vector sequence and 3' to the HER-2sv polypeptide transgene, is used to prime cDNA synthesis from the transgene transcripts. Ten mg of total spleen RNA from transgenic founders and controls is incubated with 1 mM of primer for 10 minutes at 70°C and placed on ice. The reaction is then supplemented with 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 2.5 mM $MgCl_2$, 10 mM of each dNTP, 0.1 mM DTT, and 200 U of SuperScript II reverse transcriptase. Following incubation for 50 minutes at 42°C, the reaction is stopped by heating for 15 minutes at 72°C and digested with 2U of RNase H for 20 minutes at 37°C. Samples are then amplified by PCR using primers specific for HER-2sv polypeptide.

EXAMPLE 6

Biological Activity of HER-2sv Polypeptide in Transgenic Mice

Prior to euthanasia, transgenic animals are weighed, anesthetized by isofluorane and blood drawn by cardiac puncture. The samples are subjected to hematology and serum chemistry analysis. Radiography is performed after terminal exsanguination. Upon gross dissection, major visceral organs are subject to weight analysis.

Following gross dissection, tissues (*i.e.*, liver, spleen, pancreas, stomach, the entire gastrointestinal tract, kidney, reproductive organs, skin and mammary glands, bone, brain, heart, lung, thymus, trachea, esophagus, thyroid, adrenals, urinary bladder, lymph nodes and skeletal muscle) are removed and fixed in 10% buffered Zn-Formalin for histological examination. After fixation, the tissues are processed into paraffin blocks, and 3 mm sections are obtained. All sections are stained with hematoxylin and exosin, and are then subjected to histological analysis.

The spleen, lymph node, and Peyer's patches of both the transgenic and the control mice are subjected to immunohistology analysis with B cell and T cell specific antibodies as follows. The formalin fixed paraffin embedded sections are deparaffinized and hydrated in deionized water. The sections are quenched with 3% hydrogen peroxide, blocked with Protein Block (Lipshaw, Pittsburgh, PA), and incubated in rat monoclonal anti-mouse B220 and CD3 (Harlan, Indianapolis, IN). Antibody binding is detected by biotinylated rabbit anti-rat immunoglobulins and peroxidase conjugated streptavidin (BioGenex, San Ramon, CA) with DAB as a

chromagen (BioTek, Santa Barbara, CA). Sections are counterstained with hematoxylin.

After necropsy, MLN and sections of spleen and thymus from transgenic animals and control littermates are removed. Single cell suspensions are prepared by gently grinding the tissues with the flat end of a syringe against the bottom of a 100 mm nylon cell strainer (Becton Dickinson, Franklin Lakes, NJ). Cells are washed twice, counted, and approximately 1 x 10^6 cells from each tissue are then incubated for 10 minutes with 0.5 μg CD16/32(FcγIII/II) Fc block in a 20 μL volume. Samples are then stained for 30 minutes at 2-8°C in a 100 μL volume of PBS (lacking Ca^+ and Mg^+), 0.1% bovine serum albumin, and 0.01% sodium azide with 0.5 μg antibody of FITC or PE-conjugated monoclonal antibodies against CD90.2 (Thy-1.2), CD45R (B220), CD11b (Mac-1), Gr-1, CD4, or CD8 (PharMingen, San Diego, CA). Following antibody binding, the cells are washed and then analyzed by flow cytometry on a FACScan (Becton Dickinson).

While the present invention has been described in terms of the preferred embodiments, it is understood that variations and modifications will occur to those skilled in the art. Therefore, it is intended that the appended claims cover all such equivalent variations that come within the scope of the invention as claimed.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule comprising:

(a) the nucleotide sequence as set forth in any of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9;

(b) a nucleotide sequence encoding the polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10;

(c) a nucleotide sequence that hybridizes under at least moderately stringent conditions to the complement of the nucleotide sequence of any of (a) or (b), wherein the encoded polypeptide has an activity of the polypeptide set forth in in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10; or

(d) a nucleotide sequence complementary to the nucleotide sequence of any of (a) - (c).

2. An isolated nucleic acid molecule comprising:

(a) a nucleotide sequence encoding a polypeptide that is at least about 70 percent identical to the polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, wherein the encoded polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10;

(b) a region of the nucleotide sequence of any of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9 encoding a polypeptide fragment of at least about 25 amino acid residues, wherein the polypeptide fragment has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, or is antigenic;

(c) a region of the nucleotide sequence of SEQ ID NO: 1 encoding a polypeptide fragment of at least about 25 amino acid residues, including residues 261 through 262 of SEQ ID NO: 2, wherein the polypeptide fragment has an activity of the polypeptide set forth in SEQ ID NO: 2, or is antigenic;

(d) a region of the nucleotide sequence of SEQ ID NO: 3 encoding a polypeptide fragment of at least about 25 amino acid residues, including residues 383 through 384 of SEQ ID NO: 4, wherein the polypeptide fragment has an activity of the polypeptide set forth in SEQ ID NO: 4, or is antigenic;

(e) a region of the nucleotide sequence of SEQ ID NO: 5 encoding a polypeptide fragment of at least about 25 amino acid residues, including residues 384 through 422 of SEQ ID NO: 6, wherein the polypeptide fragment has an activity of the polypeptide set forth in SEQ ID NO: 6, or is antigenic;

(f) a region of the nucleotide sequence of SEQ ID NO: 9 encoding a polypeptide fragment of at least about 25 amino acid residues, including residues 580 through 613 of SEQ ID NO: 10, wherein the polypeptide fragment has an activity of the polypeptide set forth in SEQ ID NO: 10, or is antigenic;

(g) a nucleotide sequence that hybridizes under at least moderately stringent conditions to the complement of the nucleotide sequence of any of (a) - (f), wherein the encoded polypeptide has an activity of the polypeptide set forth in in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10; or

(h) a nucleotide sequence complementary to the nucleotide sequence of any of (a) - (g).

3. An isolated nucleic acid molecule comprising:

(a) a nucleotide sequence encoding a polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10 with at least one conservative amino acid substitution, wherein the encoded polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10;

(b) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO: 2 having a C- and/or N- terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2, and wherein the polypeptide includes residues 261 through 262 of SEQ ID NO: 2;

(c) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO: 4 having a C- and/or N- terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 4, and wherein the polypeptide includes residues 383 through 384 of SEQ ID NO: 4;

(d) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO: 6 having a C- and/or N- terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 6, and wherein the polypeptide includes residues 384 through 422 of SEQ ID NO: 6;

(e) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO: 10 having a C- and/or N- terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 10, and wherein the polypeptide includes residues 580 through 613 of SEQ ID NO: 10;

(e) a nucleotide sequence encoding a polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10 with at least one modification that is an amino acid substitution, C-terminal truncation, or N-terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, and wherein the polypeptide includes residues 261 through 262 of SEQ ID NO: 2, residues 383 through 384 of SEQ ID NO: 4, residues 384 through 422 of SEQ ID NO: 6, or residues 580 through 613 of SEQ ID NO: 10;

(f) a nucleotide sequence that hybridizes under at least moderately stringent conditions to the complement of the nucleotide sequence of any of (a) - (e), wherein the encoded polypeptide has an activity of the polypeptide set forth in in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10; or

(g) a nucleotide sequence complementary to the nucleotide sequence of any of (a) - (f).

4. A vector comprising the nucleic acid molecule of any of Claims 1, 2, or 3.

5. A host cell comprising the vector of Claim 4.

6. The host cell of Claim 5 that is a eukaryotic cell.

7. The host cell of Claim 5 that is a prokaryotic cell.

8. A process of producing a HER-2sv polypeptide comprising culturing the host cell of Claim 5 under suitable conditions to express the polypeptide, and optionally isolating the polypeptide from the culture.

9. A polypeptide produced by the process of Claim 8.

10. The process of Claim 8, wherein the nucleic acid molecule comprises promoter DNA other than the promoter DNA for the native HER-2sv polypeptide operatively linked to the DNA encoding the HER-2sv polypeptide.

11. The isolated nucleic acid molecule according to Claim 2, wherein the percent identity is determined using a computer program selected from the group consisting of GAP, BLASTN, FASTA, BLASTA, BLASTX, BestFit, and the Smith-Waterman algorithm.

12. A process for determining whether a compound inhibits HER-2sv polypeptide activity or HER-2sv polypeptide production comprising exposing a cell according to any of Claims 5, 6, or 7 to the compound and measuring HER-2sv polypeptide activity or HER-2sv polypeptide production in said cell.

13. An isolated polypeptide comprising the amino acid as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10.

14. An isolated polypeptide comprising:

(a) an amino acid sequence for an ortholog of any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10;

(b) an amino acid sequence that is at least about 70 percent identical to the amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10;

(c) a fragment of the amino acid sequence set forth in SEQ ID NO: 2 comprising at least about 25 amino acid residues, including residues 261 through 262 of SEQ ID NO: 2, wherein the polypeptide fragment has an activity of the polypeptide set forth in SEQ ID NO: 2, or is antigenic;

(d) a fragment of the amino acid sequence set forth in SEQ ID NO: 4 comprising at least about 25 amino acid residues, including residues 383 through 384 of SEQ ID NO: 4, wherein the polypeptide fragment has an activity of the polypeptide set forth in SEQ ID NO: 4, or is antigenic;

(e) a fragment of the amino acid sequence set forth in SEQ ID NO: 6 comprising at least about 25 amino acid residues, including residues 384 through 422 of SEQ ID NO: 6, wherein the polypeptide fragment has an activity of the polypeptide set forth in SEQ ID NO: 6, or is antigenic; or

(f) a fragment of the amino acid sequence set forth in SEQ ID NO: 10 comprising at least about 25 amino acid residues, including residues 580 through 613 of SEQ ID NO: 10, wherein the polypeptide fragment has an activity of the polypeptide set forth in SEQ ID NO: 10, or is antigenic.

15. An isolated polypeptide comprising:

(a) the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10 with at least one conservative amino acid substitution, wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10;

(b) the amino acid sequence as set forth in SEQ ID NO: 2 having a C- and/or N- terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2, and wherein the polypeptide includes residues 261 through 262 of SEQ ID NO: 2;

(c) the amino acid sequence as set forth in SEQ ID NO: 4 having a C- and/or N- terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 4, and wherein the polypeptide includes residues 383 through 384 of SEQ ID NO: 4;

(d) the amino acid sequence as set forth in SEQ ID NO: 6 having a C- and/or N- terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 6, and wherein the polypeptide includes residues 384 through 422 of SEQ ID NO: 6;

(e) the amino acid sequence as set forth in SEQ ID NO: 10 having a C- and/or N- terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 10, and wherein the polypeptide includes residues 580 through 613 of SEQ ID NO: 10; or

(f) the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10 with at least one modification that is an amino acid substitution, C-terminal truncation, or N-terminal

truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, and wherein the polypeptide includes residues 261 through 262 of SEQ ID NO: 2, residues 383 through 384 of SEQ ID NO: 4, residues 384 through 422 of SEQ ID NO: 6, or residues 580 through 613 of SEQ ID NO: 10.

16. An isolated polypeptide encoded by the nucleic acid molecule of any of Claims 1, 2, or 3, wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10.

17. The isolated polypeptide according to Claim 14, wherein the percent identity is determined using a computer program selected from the group consisting of GAP, BLASTP, FASTA, BLASTA, BLASTX, BestFit, and the Smith-Waterman algorithm.

18. A selective binding agent or fragment thereof that specifically binds the polypeptide of any of Claims 13, 14, or 15.

19. The selective binding agent or fragment thereof of Claim 18 that specifically binds the polypeptide comprising the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, or a fragment thereof.

20. The selective binding agent of Claim 18 that is an antibody or fragment thereof.

21. The selective binding agent of Claim 18 that is a humanized antibody.

22. The selective binding agent of Claim 18 that is a human antibody or fragment thereof.

23. The selective binding agent of Claim 18 that is a polyclonal antibody or fragment thereof.

24. The selective binding agent Claim 18 that is a monoclonal antibody or fragment thereof.

25. The selective binding agent of Claim 18 that is a chimeric antibody or fragment thereof.

26. The selective binding agent of Claim 18 that is a CDR-grafted antibody or fragment thereof.

27. The selective binding agent of Claim 18 that is an antiidiotypic antibody or fragment thereof.

28. The selective binding agent of Claim 18 that is a variable region fragment.

29. The variable region fragment of Claim 28 that is a Fab or a Fab' fragment.

30. A selective binding agent or fragment thereof comprising at least one complementarity determining region with specificity for a polypeptide having the amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10.

31. The selective binding agent of Claim 18 that is bound to a detectable label.

32. The selective binding agent of Claim 18 that antagonizes HER-2sv polypeptide biological activity.

33. A method for treating, preventing, or ameliorating a HER-2sv polypeptide-related disease, condition, or disorder comprising administering to a patient an effective amount of a selective binding agent according to Claim 18.

34. A selective binding agent produced by immunizing an animal with a polypeptide comprising an amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10.

35. A hybridoma that produces a selective binding agent capable of binding a polypeptide according to any of Claims 13, 14, or 15.

36. A method of detecting or quantitating the amount of HER-2sv polypeptide using the anti-HER-2sv antibody or fragment of Claim 18.

37. A kit for detecting or quantitating the amount of HER-2sv polypeptide in a biological sample, comprising the selective binding agent of Claim 18.

38. A composition comprising the polypeptide of any of Claims 13, 14, or 15, and a pharmaceutically acceptable formulation agent.

39. The composition of Claim 38, wherein the pharmaceutically acceptable formulation agent is a carrier, adjuvant, solubilizer, stabilizer, or anti-oxidant.

40. A polypeptide comprising a derivative of the polypeptide of any of Claims 13, 14, or 15.

41. The polypeptide of Claim 40 that is covalently modified with a water-soluble polymer.

42. The polypeptide of Claim 41, wherein the water-soluble polymer is polyethylene glycol, monomethoxy-polyethylene glycol, dextran, cellulose, poly-(N-vinyl pyrrolidone) polyethylene glycol, propylene glycol homopolymers, polypropylene oxide/ethylene oxide co-polymers, polyoxyethylated polyols, or polyvinyl alcohol.

43. A composition comprising a nucleic acid molecule of any of Claims 1, 2, or 3 and a pharmaceutically acceptable formulation agent.

44. The composition of Claim 43, wherein said nucleic acid molecule is contained in a viral vector.

45. A viral vector comprising a nucleic acid molecule of any of Claims 1, 2, or 3.

46. A fusion polypeptide comprising the polypeptide of any of Claims 13, 14, or 15 fused to a heterologous amino acid sequence.

47. The fusion polypeptide of Claim 46, wherein the heterologous amino acid sequence is an IgG constant domain or fragment thereof.

48. A method for treating, preventing, or ameliorating a medical condition comprising administering to a patient the polypeptide of any of Claims 13, 14, or 15, or the polypeptide encoded by the nucleic acid of any of Claims 1, 2, or 3.

49. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or amount of expression of the polypeptide of any of Claims 13, 14, or 15, or the polypeptide encoded by the nucleic acid molecule of any of Claims 1, 2, or 3 in a sample; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or amount of expression of the polypeptide.

50. A device, comprising:

(a) a membrane suitable for implantation; and

(b) cells encapsulated within said membrane, wherein said cells secrete a protein of any of Claims 13, 14, or 15; and

said membrane is permeable to said protein and impermeable to materials detrimental to said cells.

51. A method of identifying a compound that binds to a HER-2sv polypeptide comprising:

(a) contacting the polypeptide of any of Claims 13, 14, or 15 with a compound; and

(b) determining the extent of binding of the HER-2sv polypeptide to the compound.

52. The method of Claim 51, further comprising determining the activity of the polypeptide when bound to the compound.

53. A method of modulating levels of a polypeptide in an animal comprising administering to the animal the nucleic acid molecule of any of Claims 1, 2, or 3.

54. A transgenic non-human mammal comprising the nucleic acid molecule of any of Claims 1, 2, or 3.

55. A process for determining whether a compound inhibits HER-2sv polypeptide activity or HER-2sv polypeptide production comprising exposing a transgenic mammal according to Claim 54 to the compound, and measuring HER-2sv polypeptide activity or HER-2sv polypeptide production in said mammal.

56. A nucleic acid molecule of any of Claims 1, 2, or 3 attached to a solid support.

57. An array of nucleic acid molecules comprising at least one nucleic acid molecule of any of Claims 1, 2, or 3.

FIG. 1A

```
atg gag ctg gcg gcc ttg tgc cgc tgg ggg ctc ctc ctc gcc ctc ttg   48
Met Glu Leu Ala Ala Leu Cys Arg Trp Gly Leu Leu Leu Ala Leu Leu
  1               5                  10                  15

ccc ccc gga gcc gcg agc acc caa gtg tgc acc ggc aca gac atg aag   96
Pro Pro Gly Ala Ala Ser Thr Gln Val Cys Thr Gly Thr Asp Met Lys
            20                  25                  30

ctg cgg ctc cct gcc agt ccc gag acc cac ctg gac atg ctc cgc cac   144
Leu Arg Leu Pro Ala Ser Pro Glu Thr His Leu Asp Met Leu Arg His
        35                  40                  45

ctc tac cag ggc tgc cag gtg gtg cag gga aac ctg gaa ctc acc tac   192
Leu Tyr Gln Gly Cys Gln Val Val Gln Gly Asn Leu Glu Leu Thr Tyr
    50                  55                  60

ctg ccc acc aat gcc agc ctg tcc ttc ctg cag gat atc cag gag gtg   240
Leu Pro Thr Asn Ala Ser Leu Ser Phe Leu Gln Asp Ile Gln Glu Val
65                  70                  75                  80

cag ggc tac gtg ctc atc gct cac aac caa gtg agg cag gtc cca ctg   288
Gln Gly Tyr Val Leu Ile Ala His Asn Gln Val Arg Gln Val Pro Leu
                85                  90                  95

cag agg ctg cgg att gtg cga ggc acc cag ctc ttt gag gac aac tat   336
Gln Arg Leu Arg Ile Val Arg Gly Thr Gln Leu Phe Glu Asp Asn Tyr
            100                 105                 110

gcc ctg gcc gtg cta gac aat gga gac ccg ctg aac aat acc acc cct   384
Ala Leu Ala Val Leu Asp Asn Gly Asp Pro Leu Asn Asn Thr Thr Pro
        115                 120                 125

gtc aca ggg gcc tcc cca gga ggc ctg cgg gag ctg cag ctt cga agc   432
Val Thr Gly Ala Ser Pro Gly Gly Leu Arg Glu Leu Gln Leu Arg Ser
    130                 135                 140

ctc aca gag atc ttg aaa gga ggg gtc ttg atc cag cgg aac ccc cag   480
Leu Thr Glu Ile Leu Lys Gly Gly Val Leu Ile Gln Arg Asn Pro Gln
145                 150                 155                 160

ctc tgc tac cag gac acg att ttg tgg aag gac atc ttc cac aag aac   528
Leu Cys Tyr Gln Asp Thr Ile Leu Trp Lys Asp Ile Phe His Lys Asn
                165                 170                 175

aac cag ctg gct ctc aca ctg ata gac acc aac cgc tct cgg gcc tgc   576
Asn Gln Leu Ala Leu Thr Leu Ile Asp Thr Asn Arg Ser Arg Ala Cys
            180                 185                 190

cac ccc tgt tct ctg atg tgt aag ggc tcc cgc tgc tgg gga gag agt   624
His Pro Cys Ser Leu Met Cys Lys Gly Ser Arg Cys Trp Gly Glu Ser
        195                 200                 205

tct gag gat tgt cag agc ctg acg cgc act gtc tgt gcc ggt ggc tgt   672
Ser Glu Asp Cys Gln Ser Leu Thr Arg Thr Val Cys Ala Gly Gly Cys
    210                 215                 220
```

FIG. 1B

```
gcc cgc tgc aag ggg cca ctg ccc act gac tgc tgc cat gag cag tgt   720
Ala Arg Cys Lys Gly Pro Leu Pro Thr Asp Cys Cys His Glu Gln Cys
225                 230                 235                 240

gct gcc ggc tgc acg ggc ccc aag cac tct gac tgc ctg gcc tgc ctc   768
Ala Ala Gly Cys Thr Gly Pro Lys His Ser Asp Cys Leu Ala Cys Leu
                245                 250                 255

cac ttc aac cac agt ggc atc agc tgg ctg ggg ctg cgc tca ctg agg   816
His Phe Asn His Ser Gly Ile Ser Trp Leu Gly Leu Arg Ser Leu Arg
            260                 265                 270

gaa ctg ggc agt gga ctg gcc ctc atc cac cat aac acc cac ctc tgc   864
Glu Leu Gly Ser Gly Leu Ala Leu Ile His His Asn Thr His Leu Cys
        275                 280                 285

ttc gtg cac acg gtg ccc tgg gac cag ctc ttt cgg aac ccg cac caa   912
Phe Val His Thr Val Pro Trp Asp Gln Leu Phe Arg Asn Pro His Gln
    290                 295                 300

gct ctg ctc cac act gcc aac cgg cca gag gac gag tgt gtg ggc gag   960
Ala Leu Leu His Thr Ala Asn Arg Pro Glu Asp Glu Cys Val Gly Glu
305                 310                 315                 320

ggc ctg gcc tgc cac cag ctg tgc gcc cga ggg cac tgc tgg ggt cca  1008
Gly Leu Ala Cys His Gln Leu Cys Ala Arg Gly His Cys Trp Gly Pro
                325                 330                 335

ggg ccc acc cag tgt gtc aac tgc agc cag ttc ctt cgg ggc cag gag  1056
Gly Pro Thr Gln Cys Val Asn Cys Ser Gln Phe Leu Arg Gly Gln Glu
            340                 345                 350

tgc gtg gag gaa tgc cga gta ctg cag ggg ctc ccc agg gag tat gtg  1104
Cys Val Glu Glu Cys Arg Val Leu Gln Gly Leu Pro Arg Glu Tyr Val
        355                 360                 365

aat gcc agg cac tgt ttg ccg tgc cac cct gag tgt cag ccc cag aat  1152
Asn Ala Arg His Cys Leu Pro Cys His Pro Glu Cys Gln Pro Gln Asn
    370                 375                 380

ggc tca gtg acc tgt tnn gga ccg gag gct gac cag tgt gtg gcc tgt  1200
Gly Ser Val Thr Cys Xaa Gly Pro Glu Ala Asp Gln Cys Val Ala Cys
385                 390                 395                 400

gcc cac tat aag gac cct ccc ttc tgc gtg gcc cgc tgc ccc agc ggt  1248
Ala His Tyr Lys Asp Pro Pro Phe Cys Val Ala Arg Cys Pro Ser Gly
                405                 410                 415

gtg aaa cct gac ctc tcc tac atg ccc atc tgg aag ttt cca gat gag  1296
Val Lys Pro Asp Leu Ser Tyr Met Pro Ile Trp Lys Phe Pro Asp Glu
            420                 425                 430

gag ggc gca tgc cag cct tgc ccc atc aac tgc acc cac tcc tgt gtg  1344
Glu Gly Ala Cys Gln Pro Cys Pro Ile Asn Cys Thr His Ser Cys Val
        435                 440                 445
```

FIG. 1C

```
gac ctg gat gac aag ggc tgc ccc gcc gag cag aga gcc agc cct ctg   1392
Asp Leu Asp Asp Lys Gly Cys Pro Ala Glu Gln Arg Ala Ser Pro Leu
    450                 455                 460

acg tcc atc atc tct gcg gtg gtt ggc att ctg ctg gtc gtg gtc ttg   1440
Thr Ser Ile Ile Ser Ala Val Val Gly Ile Leu Leu Val Val Val Leu
465                 470                 475                 480

ggg gtg gtc ttt ggg atc ctc atc aag cga cgg cag caa               1479
Gly Val Val Phe Gly Ile Leu Ile Lys Arg Arg Gln Gln
                485                 490
```

FIG. 2A

```
acgcgttggg agctctccca tatggtcgac ctgcaggcgg ccgcgaattc actagtgatt 60

gagccgcagt gagcacc atg gag ctg gcg gcc ttg tgc cgc tgg ggg ctc    110
                    Met Glu Leu Ala Ala Leu Cys Arg Trp Gly Leu
                      1               5                  10

ctc ctc gcc ctc ttg ccc ccc gga gcc gcg agc acc caa gtg tgc acc   158
Leu Leu Ala Leu Leu Pro Pro Gly Ala Ala Ser Thr Gln Val Cys Thr
            15                  20                  25

ggc aca gac atg aag ctg cgg ctc cct gcc agt ccc gag acc cac ctg   206
Gly Thr Asp Met Lys Leu Arg Leu Pro Ala Ser Pro Glu Thr His Leu
        30                  35                  40

gac atg ctc cgc cac ctc tac cag ggc tgc cag gtg gtg cag gga aac   254
Asp Met Leu Arg His Leu Tyr Gln Gly Cys Gln Val Val Gln Gly Asn
    45                  50                  55

ctg gaa ctc acc tac ctg ccc acc aat gcc agc ctg tcc ttc ctg cag   302
Leu Glu Leu Thr Tyr Leu Pro Thr Asn Ala Ser Leu Ser Phe Leu Gln
60                  65                  70                  75

gat atc cag gag gtg cag ggc tac gtg ctc atc gct cac aac caa gtg   350
Asp Ile Gln Glu Val Gln Gly Tyr Val Leu Ile Ala His Asn Gln Val
                80                  85                  90

agg cag gtc cca ctg cag agg ctg cgg att gtg cga ggc acc cag ctc   398
Arg Gln Val Pro Leu Gln Arg Leu Arg Ile Val Arg Gly Thr Gln Leu
            95                  100                 105

ttt gag gac aac tat gcc ctg gcc gtg cta gac aat gga gac ccg ctg   446
Phe Glu Asp Asn Tyr Ala Leu Ala Val Leu Asp Asn Gly Asp Pro Leu
        110                 115                 120

aac aat acc acc cct gtc aca ggg gcc tcc cca gga ggc ctg cgg gag   494
Asn Asn Thr Thr Pro Val Thr Gly Ala Ser Pro Gly Gly Leu Arg Glu
    125                 130                 135

ctg cag ctt cga agc ctc aca gag atc ttg aaa gga ggg gtc ttg atc   542
Leu Gln Leu Arg Ser Leu Thr Glu Ile Leu Lys Gly Gly Val Leu Ile
140                 145                 150                 155

cag cgg aac ccc cag ctc tgc tac cag gac acg att ttg tgg aag gac   590
Gln Arg Asn Pro Gln Leu Cys Tyr Gln Asp Thr Ile Leu Trp Lys Asp
                160                 165                 170

atc ttc cac aag aac aac cag ctg gct ctc aca ctg ata gac acc aac   638
Ile Phe His Lys Asn Asn Gln Leu Ala Leu Thr Leu Ile Asp Thr Asn
            175                 180                 185

cgc tct cgg gcc tgc cac ccc tgt tct ccg atg tgt aag ggc tcc cgc   686
Arg Ser Arg Ala Cys His Pro Cys Ser Pro Met Cys Lys Gly Ser Arg
        190                 195                 200

tgc tgg gga gag agt tct gag gat tgt cag agc ctg acg cgc act gtc   734
Cys Trp Gly Glu Ser Ser Glu Asp Cys Gln Ser Leu Thr Arg Thr Val
    205                 210                 215
```

FIG. 2B

```
tgt gcc ggt ggc tgt gcc cgc tgc aag ggg cca ctg ccc act gac tgc   782
Cys Ala Gly Gly Cys Ala Arg Cys Lys Gly Pro Leu Pro Thr Asp Cys
220                 225                 230                 235

tgc cat gag cag tgt gct gcc ggc tgc acg ggc ccc aag cac tct gac   830
Cys His Glu Gln Cys Ala Ala Gly Cys Thr Gly Pro Lys His Ser Asp
                240                 245                 250

tgc ctg gcc tgc ctc cac ttc aac cac agt ggc atc tgt gag ctg cac   878
Cys Leu Ala Cys Leu His Phe Asn His Ser Gly Ile Cys Glu Leu His
            255                 260                 265

tgc cca gcc ctg gtc acc tac aac aca gac acg ttt gag tcc atg ccc   926
Cys Pro Ala Leu Val Thr Tyr Asn Thr Asp Thr Phe Glu Ser Met Pro
        270                 275                 280

aat ccc gag ggc cgg tat aca ttc ggc gcc agc tgt gtg act gcc tgt   974
Asn Pro Glu Gly Arg Tyr Thr Phe Gly Ala Ser Cys Val Thr Ala Cys
    285                 290                 295

ccc tac aac tac ctt tct acg gac gtg gga tcc tgc acc ctc gtc tgc   1022
Pro Tyr Asn Tyr Leu Ser Thr Asp Val Gly Ser Cys Thr Leu Val Cys
300                 305                 310                 315

ccc ctg cac aac caa gag gtg aca gca gag gat gga aca cag cgg tgt   1070
Pro Leu His Asn Gln Glu Val Thr Ala Glu Asp Gly Thr Gln Arg Cys
                320                 325                 330

gag aag tgc agc aag ccc tgt gcc cga gtg tgc tat ggt ctg ggc atg   1118
Glu Lys Cys Ser Lys Pro Cys Ala Arg Val Cys Tyr Gly Leu Gly Met
            335                 340                 345

gag cac ttg cga gag gtg agg gca gtt acc agt gcc aat atc cag gag   1166
Glu His Leu Arg Glu Val Arg Ala Val Thr Ser Ala Asn Ile Gln Glu
        350                 355                 360

ttt gct ggc tgc aag aag atc ttt ggg agc ctg gca ttt ctg ccg gag   1214
Phe Ala Gly Cys Lys Lys Ile Phe Gly Ser Leu Ala Phe Leu Pro Glu
    365                 370                 375

agc ttt gat ggg gac cca gcc tcc aac act gcc ccg ctc cag cca gag   1262
Ser Phe Asp Gly Asp Pro Ala Ser Asn Thr Ala Pro Leu Gln Pro Glu
380                 385                 390                 395

cag ctc caa gtg ttt gag act ctg gaa gag atc aca ggt tac cta tac   1310
Gln Leu Gln Val Phe Glu Thr Leu Glu Glu Ile Thr Gly Tyr Leu Tyr
                400                 405                 410

atc tca gca tgg ccg gac agc ctg cct gac ctc agc gtc ttc cag aac   1358
Ile Ser Ala Trp Pro Asp Ser Leu Pro Asp Leu Ser Val Phe Gln Asn
            415                 420                 425

ctg caa gta atc cgg gga cga att ctg cac aat ggc gcc tac tcg ctg   1406
Leu Gln Val Ile Arg Gly Arg Ile Leu His Asn Gly Ala Tyr Ser Leu
        430                 435                 440
```

FIG. 2C

```
acc ctg caa ggg ctg ggc atc agc tgg ctg ggg ctg cgc tca ctg agg   1454
Thr Leu Gln Gly Leu Gly Ile Ser Trp Leu Gly Leu Arg Ser Leu Arg
    445                 450                 455

gaa ctg ggc agt gga ctg gcc ctc atc cac cat aac acc cac ctc tgc   1502
Glu Leu Gly Ser Gly Leu Ala Leu Ile His His Asn Thr His Leu Cys
460                 465                 470                 475

ttc gtg cac acg gtg ccc tgg gac cag ctc ttt cgg aac ccg cac caa   1550
Phe Val His Thr Val Pro Trp Asp Gln Leu Phe Arg Asn Pro His Gln
                480                 485                 490

gct ctg ctc cac act gcc aac cgg cca gag gac gag tgt gtg ggc gag   1598
Ala Leu Leu His Thr Ala Asn Arg Pro Glu Asp Glu Cys Val Gly Glu
            495                 500                 505

ggc ctg gcc tgc cac cag ctg tgc gcc cga ggg cac tgc tgg ggt cca   1646
Gly Leu Ala Cys His Gln Leu Cys Ala Arg Gly His Cys Trp Gly Pro
        510                 515                 520

ggg ccc acc cag tgt gtc aac tgc agc cag ttc ctt cgg ggc cag gag   1694
Gly Pro Thr Gln Cys Val Asn Cys Ser Gln Phe Leu Arg Gly Gln Glu
    525                 530                 535

tgc gtg gag gaa tgc cga gta ctg cag ggg ctc ccc agg gag tat gtg   1742
Cys Val Glu Glu Cys Arg Val Leu Gln Gly Leu Pro Arg Glu Tyr Val
540                 545                 550                 555

aat gcc agg cac tgt ttg ccg tgc cac cct gag tgt cag ccc cag aat   1790
Asn Ala Arg His Cys Leu Pro Cys His Pro Glu Cys Gln Pro Gln Asn
                560                 565                 570

ggc tca gtg acc tgt ttt gga ccg gag gct gac cag tgt gtg gcc tgt   1838
Gly Ser Val Thr Cys Phe Gly Pro Glu Ala Asp Gln Cys Val Ala Cys
            575                 580                 585

gcc cac tat aag gac cct ccc ttc tgc gtg gcc cgc tgc ccc agc ggt   1886
Ala His Tyr Lys Asp Pro Pro Phe Cys Val Ala Arg Cys Pro Ser Gly
        590                 595                 600

gtg aaa cct gac ctc tcc tac atg ccc atc tgg aag ttt cca gat gag   1934
Val Lys Pro Asp Leu Ser Tyr Met Pro Ile Trp Lys Phe Pro Asp Glu
    605                 610                 615

gag ggc gca tgc cag cct tgc ccc atc aac tgc acc cac tcc cct ctg   1982
Glu Gly Ala Cys Gln Pro Cys Pro Ile Asn Cys Thr His Ser Pro Leu
620                 625                 630                 635

acg tcc atc gtc tct gcg gtg gtt ggc att ctg ctg gtc gtg gtc ttg   2030
Thr Ser Ile Val Ser Ala Val Val Gly Ile Leu Leu Val Val Val Leu
                640                 645                 650

ggg gtg gtc ttt ggg atc ctc atc aag cga cgg cag caa tcg aat tcc   2078
Gly Val Val Phe Gly Ile Leu Ile Lys Arg Arg Gln Gln Ser Asn Ser
            655                 660                 665
```

FIG. 2D

```
cgc ggc cgc cat ggc ggc cgg gag cat gcg acg tcg ggc cca att cgc   2126
Arg Gly Arg His Gly Gly Arg Glu His Ala Thr Ser Gly Pro Ile Arg
        670                 675                 680

cct ata                                                           2132
Pro Ile
    685
```

FIG. 3A

```
atg gag ctg gcg gcc ttg tgc cgc tgg ggg ctc ctc ctc gcc ctc ttg   48
Met Glu Leu Ala Ala Leu Cys Arg Trp Gly Leu Leu Leu Ala Leu Leu
  1               5                  10                  15

ccc ccc gga gcc gcg agc acc caa gtg tgc acc ggc aca gac atg aag   96
Pro Pro Gly Ala Ala Ser Thr Gln Val Cys Thr Gly Thr Asp Met Lys
            20                  25                  30

ctg cgg ctc cct gcc agt ccc gag acc cac ctg gac atg ctc cgc cac  144
Leu Arg Leu Pro Ala Ser Pro Glu Thr His Leu Asp Met Leu Arg His
        35                  40                  45

ctc tac cag ggc tgc cag gtg gtg cag gga aac ctg gaa ctc acc tac  192
Leu Tyr Gln Gly Cys Gln Val Val Gln Gly Asn Leu Glu Leu Thr Tyr
    50                  55                  60

ctg ccc acc aat gcc agc ctg tcc ttc ctg cag gat atc cag gag gtg  240
Leu Pro Thr Asn Ala Ser Leu Ser Phe Leu Gln Asp Ile Gln Glu Val
65                  70                  75                  80

cag ggc tac gtg ctc atc gct cac aac caa gtg agg cag gtc cca ctg  288
Gln Gly Tyr Val Leu Ile Ala His Asn Gln Val Arg Gln Val Pro Leu
                85                  90                  95

cag agg ctg cgg att gtg cga ggc acc cag ctc ttt gag gac aac tat  336
Gln Arg Leu Arg Ile Val Arg Gly Thr Gln Leu Phe Glu Asp Asn Tyr
            100                 105                 110

gcc ctg gcc gtg cta gac aat gga gac ccg ctg aac aat acc acc cct  384
Ala Leu Ala Val Leu Asp Asn Gly Asp Pro Leu Asn Asn Thr Thr Pro
        115                 120                 125

gtc aca ggg gcc tcc cca gga ggc ctg cgg gag ctg cag ctt cga agc  432
Val Thr Gly Ala Ser Pro Gly Gly Leu Arg Glu Leu Gln Leu Arg Ser
    130                 135                 140

ctc aca gag atc ttg aaa gga ggg gtc ttg atc cag cgg aac ccc cag  480
Leu Thr Glu Ile Leu Lys Gly Gly Val Leu Ile Gln Arg Asn Pro Gln
145                 150                 155                 160

ctc tgc tac cag gac acg att ttg tgg aag gac atc ttc cac aag aac  528
Leu Cys Tyr Gln Asp Thr Ile Leu Trp Lys Asp Ile Phe His Lys Asn
                165                 170                 175

aac cag ctg gct ctc aca ctg ata gac acc aac cgc tct cgg gcc tgc  576
Asn Gln Leu Ala Leu Thr Leu Ile Asp Thr Asn Arg Ser Arg Ala Cys
            180                 185                 190

cac ccc tgt tct ccg atg tgt aag ggc tcc cgc tgc tgg gga gag agt  624
His Pro Cys Ser Pro Met Cys Lys Gly Ser Arg Cys Trp Gly Glu Ser
        195                 200                 205

tct gag gat tgt cag agc ctg acg cgc act gtc tgt gcc ggt ggc tgt  672
Ser Glu Asp Cys Gln Ser Leu Thr Arg Thr Val Cys Ala Gly Gly Cys
    210                 215                 220
```

FIG. 3B

```
gcc cgc tgc aag ggg cca ctg ccc act gac tgc tgc cat gag cag tgt   720
Ala Arg Cys Lys Gly Pro Leu Pro Thr Asp Cys Cys His Glu Gln Cys
225                 230                 235                 240

gct gcc ggc tgc acg ggc ccc aag cac tct gac tgc ctg gcc tgc ctc   768
Ala Ala Gly Cys Thr Gly Pro Lys His Ser Asp Cys Leu Ala Cys Leu
                245                 250                 255

cac ttc aac cac agt ggc atc tgt gag ctg cac tgc cca gcc ctg gtc   816
His Phe Asn His Ser Gly Ile Cys Glu Leu His Cys Pro Ala Leu Val
            260                 265                 270

acc tac aac aca gac acg ttt gag tcc atg ccc aat ccc gag ggc cgg   864
Thr Tyr Asn Thr Asp Thr Phe Glu Ser Met Pro Asn Pro Glu Gly Arg
        275                 280                 285

tat aca ttc ggc gcc agc tgt gtg act gcc tgt ccc tac aac tac ctt   912
Tyr Thr Phe Gly Ala Ser Cys Val Thr Ala Cys Pro Tyr Asn Tyr Leu
    290                 295                 300

tct acg gac gtg gga tcc tgc acc ctc gtc tgc ccc ctg cac aac caa   960
Ser Thr Asp Val Gly Ser Cys Thr Leu Val Cys Pro Leu His Asn Gln
305                 310                 315                 320

gag gtg aca gca gag gat gga aca cag cgg tgt gag aag tgc agc aag  1008
Glu Val Thr Ala Glu Asp Gly Thr Gln Arg Cys Glu Lys Cys Ser Lys
                325                 330                 335

ccc tgt gcc cga gtg tgc tat ggt ctg ggc atg gag cac ttg cga gag  1056
Pro Cys Ala Arg Val Cys Tyr Gly Leu Gly Met Glu His Leu Arg Glu
            340                 345                 350

gtg agg gca gtt acc agt gcc aat atc cag gag ttt gct ggc tgc aag  1104
Val Arg Ala Val Thr Ser Ala Asn Ile Gln Glu Phe Ala Gly Cys Lys
        355                 360                 365

aag atc ttt ggg agc ctg gca ttt ctg ccg gag agc ttt gat gga gtc  1152
Lys Ile Phe Gly Ser Leu Ala Phe Leu Pro Glu Ser Phe Asp Gly Val
    370                 375                 380

tca ctc tgt cag cag gct gga gtg cag tgg tac gat ctt ggc tca ctg  1200
Ser Leu Cys Gln Gln Ala Gly Val Gln Trp Tyr Asp Leu Gly Ser Leu
385                 390                 395                 400

caa cct ctg cct cct gga ttc aag caa ttc tcc tgc ctc agt ctc ctg  1248
Gln Pro Leu Pro Pro Gly Phe Lys Gln Phe Ser Cys Leu Ser Leu Leu
                405                 410                 415

agt agc tgg gac tac agg gac cca gcc tcc aac act gcc ccg ctc cag  1296
Ser Ser Trp Asp Tyr Arg Asp Pro Ala Ser Asn Thr Ala Pro Leu Gln
            420                 425                 430

cca gag cag ctc caa gtg ttt gag act ctg gaa gag atc aca ggt tac  1344
Pro Glu Gln Leu Gln Val Phe Glu Thr Leu Glu Glu Ile Thr Gly Tyr
        435                 440                 445
```

FIG. 3C

```
cta tac atc tca gca tgg ccg gac agc ctg cct gac ctc agc gtc ttc   1392
Leu Tyr Ile Ser Ala Trp Pro Asp Ser Leu Pro Asp Leu Ser Val Phe
    450                 455                 460

cag aac ctg caa gta atc cgg gga cga att ctg cac aat ggc gcc tac   1440
Gln Asn Leu Gln Val Ile Arg Gly Arg Ile Leu His Asn Gly Ala Tyr
465                 470                 475                 480

tcg ctg acc ctg caa ggg ctg ggc atc agc tgg ctg ggg ctg cgc tca   1488
Ser Leu Thr Leu Gln Gly Leu Gly Ile Ser Trp Leu Gly Leu Arg Ser
                485                 490                 495

ctg agg gaa ctg ggc agt gga ctg gcc ctc atc cac cat aac acc cac   1536
Leu Arg Glu Leu Gly Ser Gly Leu Ala Leu Ile His His Asn Thr His
            500                 505                 510

ctc tgc ttc gtg cac acg gtg ccc tgg gac cag ctc ttt cgg aac ccg   1584
Leu Cys Phe Val His Thr Val Pro Trp Asp Gln Leu Phe Arg Asn Pro
        515                 520                 525

cac caa gct ctg ctc cac act gcc aac cgg cca gag gac gag tgt gtg   1632
His Gln Ala Leu Leu His Thr Ala Asn Arg Pro Glu Asp Glu Cys Val
    530                 535                 540

ggc gag ggc ctg gcc tgc cac cag ctg tgc gcc cga ggg cac tgc tgg   1680
Gly Glu Gly Leu Ala Cys His Gln Leu Cys Ala Arg Gly His Cys Trp
545                 550                 555                 560

ggt cca ggg ccc acc cag tgt gtc aac tgc agc cag ttc ctt cgg ggc   1728
Gly Pro Gly Pro Thr Gln Cys Val Asn Cys Ser Gln Phe Leu Arg Gly
                565                 570                 575

cag gag tgc gtg gag gaa tgc cga gta ctg cag ggg ctc ccc agg gag   1776
Gln Glu Cys Val Glu Glu Cys Arg Val Leu Gln Gly Leu Pro Arg Glu
            580                 585                 590

tat gtg aat gcc agg cac tgt ttg ccg tgc cac cct gag tgt cag ccc   1824
Tyr Val Asn Ala Arg His Cys Leu Pro Cys His Pro Glu Cys Gln Pro
        595                 600                 605

cag aat ggc tca gtg acc tgt ttt gga ccg gag gct gac cag tgt gtg   1872
Gln Asn Gly Ser Val Thr Cys Phe Gly Pro Glu Ala Asp Gln Cys Val
    610                 615                 620

gcc tgt gcc cac tat aag gac cct ccc ttc tgc gtg gcc cgc tgc ccc   1920
Ala Cys Ala His Tyr Lys Asp Pro Pro Phe Cys Val Ala Arg Cys Pro
625                 630                 635                 640

agc ggt gtg aaa cct gac ctc tcc tac atg ccc atc tgg aag ttt cca   1968
Ser Gly Val Lys Pro Asp Leu Ser Tyr Met Pro Ile Trp Lys Phe Pro
                645                 650                 655

gat gag gag ggc gca tgc cag cct tgc ccc atc aac tgc acc cac tcc   2016
Asp Glu Glu Gly Ala Cys Gln Pro Cys Pro Ile Asn Cys Thr His Ser
            660                 665                 670
```

FIG. 3D

```
tgt gtg gac ctg gat gac aag ggc tgc ccc gcc gag cag aga gcc agc   2064
Cys Val Asp Leu Asp Asp Lys Gly Cys Pro Ala Glu Gln Arg Ala Ser
        675                 680                 685

cct ctg acg tcc atc atc tct gcg gtg gtt ggc att ctg ctg gtc gtg   2112
Pro Leu Thr Ser Ile Ile Ser Ala Val Val Gly Ile Leu Leu Val Val
    690                 695                 700

gtc ttg ggg gtg gtc ttt ggg atc ctc atc agc gac ggc agc aat cac   2160
Val Leu Gly Val Val Phe Gly Ile Leu Ile Ser Asp Gly Ser Asn His
705                 710                 715                 720

tagt                                                              2164
```

FIG. 4

```
acgcgttggg agctctccat atggtcgacc tgcaggcggc cgcgaattca ctagtgattg 60

agccgcagtg agcacc atg gag ctg gcg gcc ttg tgc cgc tgg ggg ctc ctc 112
                  Met Glu Leu Ala Ala Leu Cys Arg Trp Gly Leu Leu
                    1               5                  10

ctc gcc ctc ttg ccc ccc gga gcc gcg agc acc caa gtg tgc acc ggc   160
Leu Ala Leu Leu Pro Pro Gly Ala Ala Ser Thr Gln Val Cys Thr Gly
         15                  20                  25

aca gac atg aag ctg cgg ctc cct gcc agt ccc gag acc cac ctg gac   208
Thr Asp Met Lys Leu Arg Leu Pro Ala Ser Pro Glu Thr His Leu Asp
     30                  35                  40

atg ctc cgc cac ctc tac cag ggc tgc cag gtg gtg cag gga aac ctg   256
Met Leu Arg His Leu Tyr Gln Gly Cys Gln Val Val Gln Gly Asn Leu
 45                  50                  55                  60

gaa ctc acc tac ctg ccc acc aat gcc agc ctg tcc ttc ctg cag gat   304
Glu Leu Thr Tyr Leu Pro Thr Asn Ala Ser Leu Ser Phe Leu Gln Asp
                 65                  70                  75

atc cag gag gtg cag ggc tac gtg ctc atc gct cac aac caa gtg agg   352
Ile Gln Glu Val Gln Gly Tyr Val Leu Ile Ala His Asn Gln Val Arg
             80                  85                  90

cag gtc cca ctg cag agg ctg cgg att gtg cga ggc acc cag ctc ttt   400
Gln Val Pro Leu Gln Arg Leu Arg Ile Val Arg Gly Thr Gln Leu Phe
         95                  100                 105

gag gac aac tat gcc ctg gcc gtg cta gac aat gga gac ccg ctg aac   448
Glu Asp Asn Tyr Ala Leu Ala Val Leu Asp Asn Gly Asp Pro Leu Asn
    110                 115                 120

aat acc acc cct gtc aca ggg gcc tcc cca gga ggc ctg cgg gag ctg   496
Asn Thr Thr Pro Val Thr Gly Ala Ser Pro Gly Gly Leu Arg Glu Leu
125                 130                 135                 140

cag ctt cga agc ctc aca gag atc ttg aaa gga ggg gtc ttg atc cag   544
Gln Leu Arg Ser Leu Thr Glu Ile Leu Lys Gly Gly Val Leu Ile Gln
                145                 150                 155

cgg aac ccc cag cgg tgt gaa acc tgacctctcc tacatgccca tctggaagtt  598
Arg Asn Pro Gln Arg Cys Glu Thr
            160

tccagatgag gagggcgcat gccagccttg ccccatcaac tgcacccact cctgtgtgga 658

cctggatgac aagggctgcc ccgccgagca gagagccagc cctctgacgt ccatcatctc 718

tgcggtggtt ggcattctgc tggtcgtggt cttgggggtg gtctttggga tcctcatcaa 778

gcgacggcag caatcgaatt cccgcggccg ccatggcggc cgggagcatg cgacgtcggg 838

cccaattcgc cctatagtga gtcgtattac aattcactgg ccgtcg                 884
```

FIG. 5A

```
atg gag ctg gcg gcc ttg tgc cgc tgg ggg ctc ctc ctc gcc ctc ttg   48
Met Glu Leu Ala Ala Leu Cys Arg Trp Gly Leu Leu Leu Ala Leu Leu
  1               5                  10                  15

ccc ccc gga gcc gcg agc acc caa gtg tgc acc ggc aca gac atg aag   96
Pro Pro Gly Ala Ala Ser Thr Gln Val Cys Thr Gly Thr Asp Met Lys
            20                  25                  30

ctg cgg ctc cct gcc agt ccc gag acc cac ctg gac atg ctc cgc cac   144
Leu Arg Leu Pro Ala Ser Pro Glu Thr His Leu Asp Met Leu Arg His
        35                  40                  45

ctc tac cag ggc tgc cag gtg gtg cag gga aac ctg gaa ctc acc tac   192
Leu Tyr Gln Gly Cys Gln Val Val Gln Gly Asn Leu Glu Leu Thr Tyr
    50                  55                  60

ctg ccc acc aat gcc agc ctg tcc ttc ctg cag gat atc cag gag gtg   240
Leu Pro Thr Asn Ala Ser Leu Ser Phe Leu Gln Asp Ile Gln Glu Val
65                  70                  75                  80

cag ggc tac gtg ctc atc gct cac aac caa gtg agg cag gtc cca ctg   288
Gln Gly Tyr Val Leu Ile Ala His Asn Gln Val Arg Gln Val Pro Leu
                85                  90                  95

cag agg ctg cgg att gtg cga ggc acc cag ctc ttt gag gac aac tat   336
Gln Arg Leu Arg Ile Val Arg Gly Thr Gln Leu Phe Glu Asp Asn Tyr
            100                 105                 110

gcc ctg gcc gtg cta gac aat gga gac ccg ctg aac aat acc acc cct   384
Ala Leu Ala Val Leu Asp Asn Gly Asp Pro Leu Asn Asn Thr Thr Pro
        115                 120                 125

gtc aca ggg gcc tcc cca gga ggc ctg cgg gag ctg cag ctt cga agc   432
Val Thr Gly Ala Ser Pro Gly Gly Leu Arg Glu Leu Gln Leu Arg Ser
    130                 135                 140

ctc aca gag atc ttg aaa gga ggg gtc ttg atc cag cgg aac ccc cag   480
Leu Thr Glu Ile Leu Lys Gly Gly Val Leu Ile Gln Arg Asn Pro Gln
145                 150                 155                 160

ctc tgc tac cag gac acg att ttg tgg aag gac atc ttc cac aag aac   528
Leu Cys Tyr Gln Asp Thr Ile Leu Trp Lys Asp Ile Phe His Lys Asn
                165                 170                 175

aac cag ctg gct ctc aca ctg ata gac acc aac cgc tct cgg gcc tgc   576
Asn Gln Leu Ala Leu Thr Leu Ile Asp Thr Asn Arg Ser Arg Ala Cys
            180                 185                 190

cac ccc tgt tct ccg atg tgt aag ggc tcc cgc tgc tgg gga gag agt   624
His Pro Cys Ser Pro Met Cys Lys Gly Ser Arg Cys Trp Gly Glu Ser
        195                 200                 205

tct gag gat tgt cag agc ctg acg cgc act gtc tgt gcc ggt ggc tgt   672
Ser Glu Asp Cys Gln Ser Leu Thr Arg Thr Val Cys Ala Gly Gly Cys
    210                 215                 220
```

FIG. 5B

```
gcc cgc tgc aag ggg cca ctg ccc act gac tgc tgc cat gag cag tgt   720
Ala Arg Cys Lys Gly Pro Leu Pro Thr Asp Cys Cys His Glu Gln Cys
225                 230                 235                 240

gct gcc ggc tgc acg ggc ccc aag cac tct gac tgc ctg gcc tgc ctc   768
Ala Ala Gly Cys Thr Gly Pro Lys His Ser Asp Cys Leu Ala Cys Leu
                245                 250                 255

cac ttc aac cac agt ggc atc tgt gag ctg cac tgc cca gcc ctg gtc   816
His Phe Asn His Ser Gly Ile Cys Glu Leu His Cys Pro Ala Leu Val
            260                 265                 270

acc tac aac aca gac acg ttt gag tcc atg ccc aat ccc gag ggc cgg   864
Thr Tyr Asn Thr Asp Thr Phe Glu Ser Met Pro Asn Pro Glu Gly Arg
        275                 280                 285

tat aca ttc ggc gcc agc tgt gtg act gcc tgt ccc tac aac tac ctt   912
Tyr Thr Phe Gly Ala Ser Cys Val Thr Ala Cys Pro Tyr Asn Tyr Leu
    290                 295                 300

tct acg gac gtg gga tcc tgc acc ctc gtc tgc ccc ctg cac aac caa   960
Ser Thr Asp Val Gly Ser Cys Thr Leu Val Cys Pro Leu His Asn Gln
305                 310                 315                 320

gag gtg aca gca gag gat gga aca cag cgg tgt gag aag tgc agc aag   1008
Glu Val Thr Ala Glu Asp Gly Thr Gln Arg Cys Glu Lys Cys Ser Lys
                325                 330                 335

ccc tgt gcc cga gtg tgc tat ggt ctg ggc atg gag cac ttg cga gag   1056
Pro Cys Ala Arg Val Cys Tyr Gly Leu Gly Met Glu His Leu Arg Glu
            340                 345                 350

gtg agg gca gtt acc agt gcc aat atc cag gag ttt gct ggc tgc aag   1104
Val Arg Ala Val Thr Ser Ala Asn Ile Gln Glu Phe Ala Gly Cys Lys
        355                 360                 365

aag atc ttt ggg agc ctg gca ttt ctg ccg gag agc ttt gat ggg gac   1152
Lys Ile Phe Gly Ser Leu Ala Phe Leu Pro Glu Ser Phe Asp Gly Asp
    370                 375                 380

cca gcc tcc aac act gcc ccg ctc cag cca gag cag ctc caa gtg ttt   1200
Pro Ala Ser Asn Thr Ala Pro Leu Gln Pro Glu Gln Leu Gln Val Phe
385                 390                 395                 400

gag act ctg gaa gag atc aca ggt tac cta tac atc tca gca tgg ccg   1248
Glu Thr Leu Glu Glu Ile Thr Gly Tyr Leu Tyr Ile Ser Ala Trp Pro
                405                 410                 415

gac agc ctg cct gac ctc agc gtc ttc cag aac ctg caa gta atc cgg   1296
Asp Ser Leu Pro Asp Leu Ser Val Phe Gln Asn Leu Gln Val Ile Arg
            420                 425                 430

gga cga att ctg cac aat ggc gcc tac tcg ctg acc ctg caa ggg ctg   1344
Gly Arg Ile Leu His Asn Gly Ala Tyr Ser Leu Thr Leu Gln Gly Leu
        435                 440                 445
```

FIG. 5C

```
ggc atc agc tgg ctg ggg ctg cgc tca ctg agg gaa ctg ggc agt gga   1392
Gly Ile Ser Trp Leu Gly Leu Arg Ser Leu Arg Glu Leu Gly Ser Gly
    450                 455                 460

ctg gcc ctc atc cac cat aac acc cac ctc tgc ttc gtg cac acg gtg   1440
Leu Ala Leu Ile His His Asn Thr His Leu Cys Phe Val His Thr Val
465                 470                 475                 480

ccc tgg gac cag ctc ttt cgg aac ccg cac caa gct ctg ctc cac act   1488
Pro Trp Asp Gln Leu Phe Arg Asn Pro His Gln Ala Leu Leu His Thr
                485                 490                 495

gcc aac cgg cca gag gac gag tgt gtg ggc gag ggc ctg gcc tgc cac   1536
Ala Asn Arg Pro Glu Asp Glu Cys Val Gly Glu Gly Leu Ala Cys His
            500                 505                 510

cag ctg tgc gcc cga ggg cac tgc tgg ggt cca ggg ccc acc cag tgt   1584
Gln Leu Cys Ala Arg Gly His Cys Trp Gly Pro Gly Pro Thr Gln Cys
        515                 520                 525

gtc aac tgc agc cag ttc ctt cgg ggc cag gag tgc gtg gag gaa tgc   1632
Val Asn Cys Ser Gln Phe Leu Arg Gly Gln Glu Cys Val Glu Glu Cys
    530                 535                 540

cga gta ctg cag ggg ctc ccc agg gag tat gtg aat gcc agg cac tgt   1680
Arg Val Leu Gln Gly Leu Pro Arg Glu Tyr Val Asn Ala Arg His Cys
545                 550                 555                 560

ttg ccg tgc cac cct gag tgt cag ccc cag aat ggc tca gtg acc tgt   1728
Leu Pro Cys His Pro Glu Cys Gln Pro Gln Asn Gly Ser Val Thr Cys
                565                 570                 575

ttt gga ccg gta atg cgt ttt cct ctc tgg gtg cct ccc att ttc tgg   1776
Phe Gly Pro Val Met Arg Phe Pro Leu Trp Val Pro Pro Ile Phe Trp
            580                 585                 590

ctc aag tcc ctg ccc agg atc aag ctt gga gga ggg ccc cga ggg agg   1824
Leu Lys Ser Leu Pro Arg Ile Lys Leu Gly Gly Gly Pro Arg Gly Arg
        595                 600                 605

ggc cac aga gac tgg gag gct gac cag tgt gtg gcc tgt gcc cac tat   1872
Gly His Arg Asp Trp Glu Ala Asp Gln Cys Val Ala Cys Ala His Tyr
    610                 615                 620

aag gac cct ccc ttc tgc gtg gcc cga tgc ccc agc ggt gtg aaa cct   1920
Lys Asp Pro Pro Phe Cys Val Ala Arg Cys Pro Ser Gly Val Lys Pro
625                 630                 635                 640

gac ctc tcc tac atg ccc atc tgg aag ttt cca gat gag gag ggc gca   1968
Asp Leu Ser Tyr Met Pro Ile Trp Lys Phe Pro Asp Glu Glu Gly Ala
                645                 650                 655

tgc cag cct tgc ccc atc aac tgc acc cac tcc tgt gtg gac ctg gat   2016
Cys Gln Pro Cys Pro Ile Asn Cys Thr His Ser Cys Val Asp Leu Asp
            660                 665                 670
```

FIG. 5D

```
gac aag ggc tgc ccc gcc gag cag aga gcc agc cct ctg atg tcc atc   2064
Asp Lys Gly Cys Pro Ala Glu Gln Arg Ala Ser Pro Leu Met Ser Ile
        675                 680                 685

atc tct gcg gtg gtt ggc att ctg ctg gtc gtg gtc ttg ggg gtg gtc   2112
Ile Ser Ala Val Val Gly Ile Leu Leu Val Val Val Leu Gly Val Val
    690                 695                 700

ttt ggg atc ctc ata agc gac ggc agc aat cac tagt                 2149
Phe Gly Ile Leu Ile Ser Asp Gly Ser Asn His
705                 710                 715
```

FIG. 6A

```
          1                                                        50
   HER-2 MELAALCRWG LLLALLPPGA ASTQVCTGTD MKLRLPASPE THLDMLRHLY
 form 97 MELAALCRWG LLLALLPPGA ASTQVCTGTD MKLRLPASPE THLDMLRHLY
form 184 MELAALCRWG LLLALLPPGA ASTQVCTGTD MKLRLPASPE THLDMLRHLY
form 119 MELAALCRWG LLLALLPPGA ASTQVCTGTD MKLRLPASPE THLDMLRHLY
 form 68 MELAALCRWG LLLALLPPGA ASTQVCTGTD MKLRLPASPE THLDMLRHLY
form 156 MELAALCRWG LLLALLPPGA ASTQVCTGTD MKLRLPASPE THLDMLRHLY

          51                                                      100
   HER-2 QGCQVVQGNL ELTYLPTNAS LSFLQDIQEV QGYVLIAHNQ VRQVPLQRLR
 form 97 QGCQVVQGNL ELTYLPTNAS LSFLQDIQEV QGYVLIAHNQ VRQVPLQRLR
form 184 QGCQVVQGNL ELTYLPTNAS LSFLQDIQEV QGYVLIAHNQ VRQVPLQRLR
form 119 QGCQVVQGNL ELTYLPTNAS LSFLQDIQEV QGYVLIAHNQ VRQVPLQRLR
 form 68 QGCQVVQGNL ELTYLPTNAS LSFLQDIQEV QGYVLIAHNQ VRQVPLQRLR
form 156 QGCQVVQGNL ELTYLPTNAS LSFLQDIQEV QGYVLIAHNQ VRQVPLQRLR

          101                                                     150
   HER-2 IVRGTQLFED NYALAVLDNG DPLNNTTPVT GASPGGLREL QLRSLTEILK
 form 97 IVRGTQLFED NYALAVLDNG DPLNNTTPVT GASPGGLREL QLRSLTEILK
form 184 IVRGTQLFED NYALAVLDNG DPLNNTTPVT GASPGGLREL QLRSLTEILK
form 119 IVRGTQLFED NYALAVLDNG DPLNNTTPVT GASPGGLREL QLRSLTEILK
 form 68 IVRGTQLFED NYALAVLDNG DPLNNTTPVT GASPGGLREL QLRSLTEILK
form 156 IVRGTQLFED NYALAVLDNG DPLNNTTPVT GASPGGLREL QLRSLTEILK

          151                                                     200
   HER-2 GGVLIQRNPQ LCYQDTILWK DIFHKNNQLA LTLIDTNRSR ACHPCSPMCK
 form 97 GGVLIQRNPQ LCYQDTILWK DIFHKNNQLA LTLIDTNRSR ACHPCSPMCK
form 184 GGVLIQRNPQ LCYQDTILWK DIFHKNNQLA LTLIDTNRSR ACHPCSPMCK
form 119 GGVLIQRNPQ LCYQDTILWK DIFHKNNQLA LTLIDTNRSR ACHPCSPMCK
 form 68 GGVLIQRNPQ LCYQDTILWK DIFHKNNQLA LTLIDTNRSR ACHPCSLMCK
form 156 GGVLIQRNPQ RCET*~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

          201                                                     250
   HER-2 GSRCWGESSE DCQSLTRTVC AGGCARCKGP LPTDCCHEQC AAGCTGPKHS
 form 97 GSRCWGESSE DCQSLTRTVC AGGCARCKGP LPTDCCHEQC AAGCTGPKHS
form 184 GSRCWGESSE DCQSLTRTVC AGGCARCKGP LPTDCCHEQC AAGCTGPKHS
form 119 GSRCWGESSE DCQSLTRTVC AGGCARCKGP LPTDCCHEQC AAGCTGPKHS
 form 68 GSRCWGESSE DCQSLTRTVC AGGCARCKGP LPTDCCHEQC AAGCTGPKHS
form 156 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

          251                                                     300
   HER-2 DCLACLHFNH SGICELHCPA LVTYNTDTFE SMPNPEGRYT FGASCVTACP
 form 97 DCLACLHFNH SGICELHCPA LVTYNTDTFE SMPNPEGRYT FGASCVTACP
form 184 DCLACLHFNH SGICELHCPA LVTYNTDTFE SMPNPEGRYT FGASCVTACP
form 119 DCLACLHFNH SGICELHCPA LVTYNTDTFE SMPNPEGRYT FGASCVTACP
 form 68 DCLACLHFNH S......... .......... .......... ..........
form 156 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
```

FIG. 6B

```
        301                                                        350
   HER-2 YNYLSTDVGS CTLVCPLHNQ EVTAEDGTQR CEKCSKPCAR VCYGLGMEHL
 form 97 YNYLSTDVGS CTLVCPLHNQ EVTAEDGTQR CEKCSKPCAR VCYGLGMEHL
form 184 YNYLSTDVGS CTLVCPLHNQ EVTAEDGTQR CEKCSKPCAR VCYGLGMEHL
form 119 YNYLSTDVGS CTLVCPLHNQ EVTAEDGTQR CEKCSKPCAR VCYGLGMEHL
 form 68 .......... .......... .......... .......... ..........
form 156 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

        351                                                        400
   HER-2 REVRAVTSAN IQEFAGCKKI FGSLAFLPES FDG....... ..........
 form 97 REVRAVTSAN IQEFAGCKKI FGSLAFLPES FDG....... ..........
form 184 REVRAVTSAN IQEFAGCKKI FGSLAFLPES FDG....... ..........
form 119 REVRAVTSAN IQEFAGCKKI FGSLAFLPES FDGVSLCQQA GVQWYDLGSL
 form 68 .......... .......... .......... .......... ..........
form 156 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

        401                                                        450
   HER-2 .......... .......... ..DPASNTAP LQPEQLQVFE TLEEITGYLY
 form 97 .......... .......... ..DPASNTAP LQPEQLQVFE TLEEITGYLY
form 184 .......... .......... ..DPASNTAP LQPEQLQVFE TLEEITGYLY
form 119 QPLPPGFKQF SCLSLLSSWD YRDPASNTAP LQPEQLQVFE TLEEITGYLY
 form 68 .......... .......... .......... .......... ..........
form 156 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

        451                                                        500
   HER-2 ISAWPDSLPD LSVFQNLQVI RGRILHNGAY SLTLQGLGIS WLGLRSLREL
 form 97 ISAWPDSLPD LSVFQNLQVI RGRILHNGAY SLTLQGLGIS WLGLRSLREL
form 184 ISAWPDSLPD LSVFQNLQVI RGRILHNGAY SLTLQGLGIS WLGLRSLREL
form 119 ISAWPDSLPD LSVFQNLQVI RGRILHNGAY SLTLQGLGIS WLGLRSLREL
 form 68 .......... .......... .......... .......GIS WLGLRSLREL
form 156 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

        501                                                        550
   HER-2 GSGLALIHHN THLCFVHTVP WDQLFRNPHQ ALLHTANRPE DECVGEGLAC
 form 97 GSGLALIHHN THLCFVHTVP WDQLFRNPHQ ALLHTANRPE DECVGEGLAC
form 184 GSGLALIHHN THLCFVHTVP WDQLFRNPHQ ALLHTANRPE DECVGEGLAC
form 119 GSGLALIHHN THLCFVHTVP WDQLFRNPHQ ALLHTANRPE DECVGEGLAC
 form 68 GSGLALIHHN THLCFVHTVP WDQLFRNPHQ ALLHTANRPE DECVGEGLAC
form 156 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

        551                                                        600
   HER-2 HQLCARGHCW GPGPTQCVNC SQFLRGQECV EECRVLQGLP REYVNARHCL
 form 97 HQLCARGHCW GPGPTQCVNC SQFLRGQECV EECRVLQGLP REYVNARHCL
form 184 HQLCARGHCW GPGPTQCVNC SQFLRGQECV EECRVLQGLP REYVNARHCL
form 119 HQLCARGHCW GPGPTQCVNC SQFLRGQECV EECRVLQGLP REYVNARHCL
 form 68 HQLCARGHCW GPGPTQCVNC SQFLRGQECV EECRVLQGLP REYVNARHCL
form 156 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
```

FIG. 6C

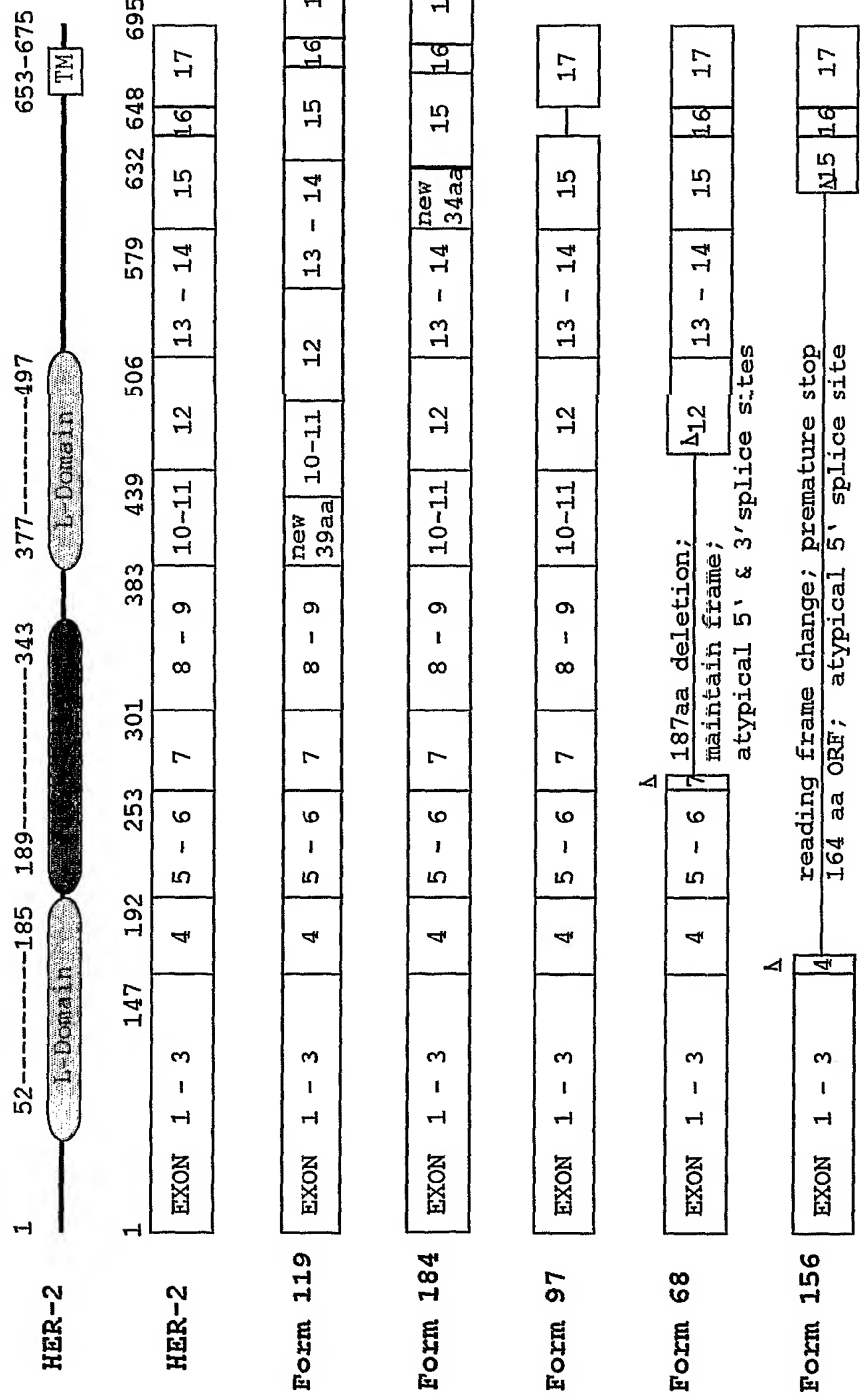
```
        601                                                        650
   HER-2 PCHPECQPQN GSVTCFGP.. .......... .......... ..........
 form 97 PCHPECQPQN GSVTCFGP.. .......... .......... ..........
form 184 PCHPECQPQN GSVTCFGPVM RFPLWVPPIF WLKSLPRIKL GGGPRGRGHR
form 119 PCHPECQPQN GSVTCFGP.. .......... .......... ..........
 form 68 PCHPECQPQN GSVTCXGP.. .......... .......... ..........
form 156 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

        651                                                        700
   HER-2 ..EADQCVAC AHYKDPPFCV ARCPSGVKPD LSYMPIWKFP DEEGACQPCP
 form 97 ..EADQCVAC AHYKDPPFCV ARCPSGVKPD LSYMPIWKFP DEEGACQPCP
form 184 DWEADQCVAC AHYKDPPFCV ARCPSGVKPD LSYMPIWKFP DEEGACQPCP
form 119 ..EADQCVAC AHYKDPPFCV ARCPSGVKPD LSYMPIWKFP DEEGACQPCP
 form 68 ..EADQCVAC AHYKDPPFCV ARCPSGVKPD LSYMPIWKFP DEEGACQPCP
form 156 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

        701                                                        750
   HER-2 INCTHSCVDL DDKGCPAEQR ASPLTSIVSA VVGILLVVVL GVVFGILIKR
 form 97 INCTH..... .......... .SPLTSIVSA VVGILLVVVL GVVFGILIKR
form 184 INCTHSCVDL DDKGCPAEQR ASPLMSIISA VVGILLVVVL GVVFGILISD
form 119 INCTHSCVDL DDKGCPAEQR ASPLTSIISA VVGILLVVVL GVVFGILISD
 form 68 INCTHSCVDL DDKGCPAEQR ASPLTSIISA VVGILLVVVL GVVFGILIKR
form 156 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

        751          763
   HER-2 RQQKIRKYTM RRL
 form 97 RQQ~~~~~~~ ~~~
form 184 GSNH~~~~~~ ~~~
form 119 GSNH~~~~~~ ~~~
 form 68 RQQ~~~~~~~ ~~~
form 156 ~~~~~~~~~~ ~~~
```

FIG. 7


HER-2
1
52
185
189
343
377
497
653-675
L-Domain
L-Domain
TM
HER-2
1
147
192
253
301
383
439
506
579
632
648
695
EXON 1 - 3
4
5 - 6
7
8 - 9
10-11
12
13 - 14
15
16
17
Form 119
new 39aa
Form 184
new 34aa
Form 97
Form 68
Δ
187aa deletion;
maintain frame;
atypical 5' & 3' splice sites
Δ12
Form 156
reading frame change; premature stop
164 aa ORF; atypical 5' splice site
Δ15

SEQUENCE LISTING

<110> Jing, Shuqian
Tatarewicz, Suzanna

<120> HER-2 Receptor Tyrosine Kinase Molecules and Uses Thereof

<130> 01-1624-B

<140>
<141>

<160> 17

<170> PatentIn Ver. 2.0

<210> 1
<211> 1479
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(1479)

<400> 1

```
atg gag ctg gcg gcc ttg tgc cgc tgg ggg ctc ctc ctc gcc ctc ttg   48
Met Glu Leu Ala Ala Leu Cys Arg Trp Gly Leu Leu Leu Ala Leu Leu
  1               5                  10                  15

ccc ccc gga gcc gcg agc acc caa gtg tgc acc ggc aca gac atg aag   96
Pro Pro Gly Ala Ala Ser Thr Gln Val Cys Thr Gly Thr Asp Met Lys
            20                  25                  30

ctg cgg ctc cct gcc agt ccc gag acc cac ctg gac atg ctc cgc cac   144
Leu Arg Leu Pro Ala Ser Pro Glu Thr His Leu Asp Met Leu Arg His
        35                  40                  45

ctc tac cag ggc tgc cag gtg gtg cag gga aac ctg gaa ctc acc tac   192
Leu Tyr Gln Gly Cys Gln Val Val Gln Gly Asn Leu Glu Leu Thr Tyr
    50                  55                  60

ctg ccc acc aat gcc agc ctg tcc ttc ctg cag gat atc cag gag gtg   240
Leu Pro Thr Asn Ala Ser Leu Ser Phe Leu Gln Asp Ile Gln Glu Val
65                  70                  75                  80

cag ggc tac gtg ctc atc gct cac aac caa gtg agg cag gtc cca ctg   288
Gln Gly Tyr Val Leu Ile Ala His Asn Gln Val Arg Gln Val Pro Leu
                85                  90                  95

cag agg ctg cgg att gtg cga ggc acc cag ctc ttt gag gac aac tat   336
Gln Arg Leu Arg Ile Val Arg Gly Thr Gln Leu Phe Glu Asp Asn Tyr
            100                 105                 110

gcc ctg gcc gtg cta gac aat gga gac ccg ctg aac aat acc acc cct   384
Ala Leu Ala Val Leu Asp Asn Gly Asp Pro Leu Asn Asn Thr Thr Pro
        115                 120                 125

gtc aca ggg gcc tcc cca gga ggc ctg cgg gag ctg cag ctt cga agc   432
Val Thr Gly Ala Ser Pro Gly Gly Leu Arg Glu Leu Gln Leu Arg Ser
```

```
    130                 135                 140

ctc aca gag atc ttg aaa gga ggg gtc ttg atc cag cgg aac ccc cag   480
Leu Thr Glu Ile Leu Lys Gly Gly Val Leu Ile Gln Arg Asn Pro Gln
145                 150                 155                 160

ctc tgc tac cag gac acg att ttg tgg aag gac atc ttc cac aag aac   528
Leu Cys Tyr Gln Asp Thr Ile Leu Trp Lys Asp Ile Phe His Lys Asn
                165                 170                 175

aac cag ctg gct ctc aca ctg ata gac acc aac cgc tct cgg gcc tgc   576
Asn Gln Leu Ala Leu Thr Leu Ile Asp Thr Asn Arg Ser Arg Ala Cys
            180                 185                 190

cac ccc tgt tct ctg atg tgt aag ggc tcc cgc tgc tgg gga gag agt   624
His Pro Cys Ser Leu Met Cys Lys Gly Ser Arg Cys Trp Gly Glu Ser
        195                 200                 205

tct gag gat tgt cag agc ctg acg cgc act gtc tgt gcc ggt ggc tgt   672
Ser Glu Asp Cys Gln Ser Leu Thr Arg Thr Val Cys Ala Gly Gly Cys
    210                 215                 220

gcc cgc tgc aag ggg cca ctg ccc act gac tgc tgc cat gag cag tgt   720
Ala Arg Cys Lys Gly Pro Leu Pro Thr Asp Cys Cys His Glu Gln Cys
225                 230                 235                 240

gct gcc ggc tgc acg ggc ccc aag cac tct gac tgc ctg gcc tgc ctc   768
Ala Ala Gly Cys Thr Gly Pro Lys His Ser Asp Cys Leu Ala Cys Leu
                245                 250                 255

cac ttc aac cac agt ggc atc agc tgg ctg ggg ctg cgc tca ctg agg   816
His Phe Asn His Ser Gly Ile Ser Trp Leu Gly Leu Arg Ser Leu Arg
            260                 265                 270

gaa ctg ggc agt gga ctg gcc ctc atc cac cat aac acc cac ctc tgc   864
Glu Leu Gly Ser Gly Leu Ala Leu Ile His His Asn Thr His Leu Cys
        275                 280                 285

ttc gtg cac acg gtg ccc tgg gac cag ctc ttt cgg aac ccg cac caa   912
Phe Val His Thr Val Pro Trp Asp Gln Leu Phe Arg Asn Pro His Gln
    290                 295                 300

gct ctg ctc cac act gcc aac cgg cca gag gac gag tgt gtg ggc gag   960
Ala Leu Leu His Thr Ala Asn Arg Pro Glu Asp Glu Cys Val Gly Glu
305                 310                 315                 320

ggc ctg gcc tgc cac cag ctg tgc gcc cga ggg cac tgc tgg ggt cca  1008
Gly Leu Ala Cys His Gln Leu Cys Ala Arg Gly His Cys Trp Gly Pro
                325                 330                 335

ggg ccc acc cag tgt gtc aac tgc agc cag ttc ctt cgg ggc cag gag  1056
Gly Pro Thr Gln Cys Val Asn Cys Ser Gln Phe Leu Arg Gly Gln Glu
            340                 345                 350

tgc gtg gag gaa tgc cga gta ctg cag ggg ctc ccc agg gag tat gtg  1104
Cys Val Glu Glu Cys Arg Val Leu Gln Gly Leu Pro Arg Glu Tyr Val
        355                 360                 365

aat gcc agg cac tgt ttg ccg tgc cac cct gag tgt cag ccc cag aat  1152
Asn Ala Arg His Cys Leu Pro Cys His Pro Glu Cys Gln Pro Gln Asn
    370                 375                 380
```

```
ggc tca gtg acc tgt tnn gga ccg gag gct gac cag tgt gtg gcc tgt   1200
Gly Ser Val Thr Cys Xaa Gly Pro Glu Ala Asp Gln Cys Val Ala Cys
385                 390                 395                 400

gcc cac tat aag gac cct ccc ttc tgc gtg gcc cgc tgc ccc agc ggt   1248
Ala His Tyr Lys Asp Pro Pro Phe Cys Val Ala Arg Cys Pro Ser Gly
                405                 410                 415

gtg aaa cct gac ctc tcc tac atg ccc atc tgg aag ttt cca gat gag   1296
Val Lys Pro Asp Leu Ser Tyr Met Pro Ile Trp Lys Phe Pro Asp Glu
            420                 425                 430

gag ggc gca tgc cag cct tgc ccc atc aac tgc acc cac tcc tgt gtg   1344
Glu Gly Ala Cys Gln Pro Cys Pro Ile Asn Cys Thr His Ser Cys Val
        435                 440                 445

gac ctg gat gac aag ggc tgc ccc gcc gag cag aga gcc agc cct ctg   1392
Asp Leu Asp Asp Lys Gly Cys Pro Ala Glu Gln Arg Ala Ser Pro Leu
    450                 455                 460

acg tcc atc atc tct gcg gtg gtt ggc att ctg ctg gtc gtg gtc ttg   1440
Thr Ser Ile Ile Ser Ala Val Val Gly Ile Leu Leu Val Val Val Leu
465                 470                 475                 480

ggg gtg gtc ttt ggg atc ctc atc aag cga cgg cag caa               1479
Gly Val Val Phe Gly Ile Leu Ile Lys Arg Arg Gln Gln
                485                 490
```

<210> 2
<211> 493
<212> PRT
<213> Homo sapiens

<400> 2

```
Met Glu Leu Ala Ala Leu Cys Arg Trp Gly Leu Leu Leu Ala Leu Leu
  1               5                  10                  15

Pro Pro Gly Ala Ala Ser Thr Gln Val Cys Thr Gly Thr Asp Met Lys
            20                  25                  30

Leu Arg Leu Pro Ala Ser Pro Glu Thr His Leu Asp Met Leu Arg His
        35                  40                  45

Leu Tyr Gln Gly Cys Gln Val Val Gln Gly Asn Leu Glu Leu Thr Tyr
    50                  55                  60

Leu Pro Thr Asn Ala Ser Leu Ser Phe Leu Gln Asp Ile Gln Glu Val
65                  70                  75                  80

Gln Gly Tyr Val Leu Ile Ala His Asn Gln Val Arg Gln Val Pro Leu
                85                  90                  95

Gln Arg Leu Arg Ile Val Arg Gly Thr Gln Leu Phe Glu Asp Asn Tyr
            100                 105                 110

Ala Leu Ala Val Leu Asp Asn Gly Asp Pro Leu Asn Asn Thr Thr Pro
        115                 120                 125

Val Thr Gly Ala Ser Pro Gly Gly Leu Arg Glu Leu Gln Leu Arg Ser
```

130 135 140

Leu Thr Glu Ile Leu Lys Gly Gly Val Leu Ile Gln Arg Asn Pro Gln
145 150 155 160

Leu Cys Tyr Gln Asp Thr Ile Leu Trp Lys Asp Ile Phe His Lys Asn
165 170 175

Asn Gln Leu Ala Leu Thr Leu Ile Asp Thr Asn Arg Ser Arg Ala Cys
180 185 190

His Pro Cys Ser Leu Met Cys Lys Gly Ser Arg Cys Trp Gly Glu Ser
195 200 205

Ser Glu Asp Cys Gln Ser Leu Thr Arg Thr Val Cys Ala Gly Gly Cys
210 215 220

Ala Arg Cys Lys Gly Pro Leu Pro Thr Asp Cys Cys His Glu Gln Cys
225 230 235 240

Ala Ala Gly Cys Thr Gly Pro Lys His Ser Asp Cys Leu Ala Cys Leu
245 250 255

His Phe Asn His Ser Gly Ile Ser Trp Leu Gly Leu Arg Ser Leu Arg
260 265 270

Glu Leu Gly Ser Gly Leu Ala Leu Ile His His Asn Thr His Leu Cys
275 280 285

Phe Val His Thr Val Pro Trp Asp Gln Leu Phe Arg Asn Pro His Gln
290 295 300

Ala Leu Leu His Thr Ala Asn Arg Pro Glu Asp Glu Cys Val Gly Glu
305 310 315 320

Gly Leu Ala Cys His Gln Leu Cys Ala Arg Gly His Cys Trp Gly Pro
325 330 335

Gly Pro Thr Gln Cys Val Asn Cys Ser Gln Phe Leu Arg Gly Gln Glu
340 345 350

Cys Val Glu Glu Cys Arg Val Leu Gln Gly Leu Pro Arg Glu Tyr Val
355 360 365

Asn Ala Arg His Cys Leu Pro Cys His Pro Glu Cys Gln Pro Gln Asn
370 375 380

Gly Ser Val Thr Cys Xaa Gly Pro Glu Ala Asp Gln Cys Val Ala Cys
385 390 395 400

Ala His Tyr Lys Asp Pro Pro Phe Cys Val Ala Arg Cys Pro Ser Gly
405 410 415

Val Lys Pro Asp Leu Ser Tyr Met Pro Ile Trp Lys Phe Pro Asp Glu
420 425 430

Glu Gly Ala Cys Gln Pro Cys Pro Ile Asn Cys Thr His Ser Cys Val
435 440 445

Asp Leu Asp Asp Lys Gly Cys Pro Ala Glu Gln Arg Ala Ser Pro Leu
450 455 460

```
Thr Ser Ile Ile Ser Ala Val Val Gly Ile Leu Leu Val Val Val Leu
465                 470                 475                 480

Gly Val Val Phe Gly Ile Leu Ile Lys Arg Arg Gln Gln
                485                 490
```

<210> 3
<211> 2132
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (78)..(2132)

<400> 3
acgcgttggg agctctccca tatggtcgac ctgcaggcgg ccgcgaattc actagtgatt 60

```
gagccgcagt gagcacc atg gag ctg gcg gcc ttg tgc cgc tgg ggg ctc    110
                    Met Glu Leu Ala Ala Leu Cys Arg Trp Gly Leu
                      1               5                  10

ctc ctc gcc ctc ttg ccc ccc gga gcc gcg agc acc caa gtg tgc acc   158
Leu Leu Ala Leu Leu Pro Pro Gly Ala Ala Ser Thr Gln Val Cys Thr
            15                  20                  25

ggc aca gac atg aag ctg cgg ctc cct gcc agt ccc gag acc cac ctg   206
Gly Thr Asp Met Lys Leu Arg Leu Pro Ala Ser Pro Glu Thr His Leu
        30                  35                  40

gac atg ctc cgc cac ctc tac cag ggc tgc cag gtg gtg cag gga aac   254
Asp Met Leu Arg His Leu Tyr Gln Gly Cys Gln Val Val Gln Gly Asn
    45                  50                  55

ctg gaa ctc acc tac ctg ccc acc aat gcc agc ctg tcc ttc ctg cag   302
Leu Glu Leu Thr Tyr Leu Pro Thr Asn Ala Ser Leu Ser Phe Leu Gln
60                  65                  70                  75

gat atc cag gag gtg cag ggc tac gtg ctc atc gct cac aac caa gtg   350
Asp Ile Gln Glu Val Gln Gly Tyr Val Leu Ile Ala His Asn Gln Val
                80                  85                  90

agg cag gtc cca ctg cag agg ctg cgg att gtg cga ggc acc cag ctc   398
Arg Gln Val Pro Leu Gln Arg Leu Arg Ile Val Arg Gly Thr Gln Leu
            95                  100                 105

ttt gag gac aac tat gcc ctg gcc gtg cta gac aat gga gac ccg ctg   446
Phe Glu Asp Asn Tyr Ala Leu Ala Val Leu Asp Asn Gly Asp Pro Leu
        110                 115                 120

aac aat acc acc cct gtc aca ggg gcc tcc cca gga ggc ctg cgg gag   494
Asn Asn Thr Thr Pro Val Thr Gly Ala Ser Pro Gly Gly Leu Arg Glu
    125                 130                 135

ctg cag ctt cga agc ctc aca gag atc ttg aaa gga ggg gtc ttg atc   542
Leu Gln Leu Arg Ser Leu Thr Glu Ile Leu Lys Gly Gly Val Leu Ile
140                 145                 150                 155

cag cgg aac ccc cag ctc tgc tac cag gac acg att ttg tgg aag gac   590
```

```
Gln Arg Asn Pro Gln Leu Cys Tyr Gln Asp Thr Ile Leu Trp Lys Asp
                160                 165                 170

atc ttc cac aag aac aac cag ctg gct ctc aca ctg ata gac acc aac   638
Ile Phe His Lys Asn Asn Gln Leu Ala Leu Thr Leu Ile Asp Thr Asn
            175                 180                 185

cgc tct cgg gcc tgc cac ccc tgt tct ccg atg tgt aag ggc tcc cgc   686
Arg Ser Arg Ala Cys His Pro Cys Ser Pro Met Cys Lys Gly Ser Arg
        190                 195                 200

tgc tgg gga gag agt tct gag gat tgt cag agc ctg acg cgc act gtc   734
Cys Trp Gly Glu Ser Ser Glu Asp Cys Gln Ser Leu Thr Arg Thr Val
    205                 210                 215

tgt gcc ggt ggc tgt gcc cgc tgc aag ggg cca ctg ccc act gac tgc   782
Cys Ala Gly Gly Cys Ala Arg Cys Lys Gly Pro Leu Pro Thr Asp Cys
220                 225                 230                 235

tgc cat gag cag tgt gct gcc ggc tgc acg ggc ccc aag cac tct gac   830
Cys His Glu Gln Cys Ala Ala Gly Cys Thr Gly Pro Lys His Ser Asp
                240                 245                 250

tgc ctg gcc tgc ctc cac ttc aac cac agt ggc atc tgt gag ctg cac   878
Cys Leu Ala Cys Leu His Phe Asn His Ser Gly Ile Cys Glu Leu His
            255                 260                 265

tgc cca gcc ctg gtc acc tac aac aca gac acg ttt gag tcc atg ccc   926
Cys Pro Ala Leu Val Thr Tyr Asn Thr Asp Thr Phe Glu Ser Met Pro
        270                 275                 280

aat ccc gag ggc cgg tat aca ttc ggc gcc agc tgt gtg act gcc tgt   974
Asn Pro Glu Gly Arg Tyr Thr Phe Gly Ala Ser Cys Val Thr Ala Cys
    285                 290                 295

ccc tac aac tac ctt tct acg gac gtg gga tcc tgc acc ctc gtc tgc   1022
Pro Tyr Asn Tyr Leu Ser Thr Asp Val Gly Ser Cys Thr Leu Val Cys
300                 305                 310                 315

ccc ctg cac aac caa gag gtg aca gca gag gat gga aca cag cgg tgt   1070
Pro Leu His Asn Gln Glu Val Thr Ala Glu Asp Gly Thr Gln Arg Cys
                320                 325                 330

gag aag tgc agc aag ccc tgt gcc cga gtg tgc tat ggt ctg ggc atg   1118
Glu Lys Cys Ser Lys Pro Cys Ala Arg Val Cys Tyr Gly Leu Gly Met
            335                 340                 345

gag cac ttg cga gag gtg agg gca gtt acc agt gcc aat atc cag gag   1166
Glu His Leu Arg Glu Val Arg Ala Val Thr Ser Ala Asn Ile Gln Glu
        350                 355                 360

ttt gct ggc tgc aag aag atc ttt ggg agc ctg gca ttt ctg ccg gag   1214
Phe Ala Gly Cys Lys Lys Ile Phe Gly Ser Leu Ala Phe Leu Pro Glu
    365                 370                 375

agc ttt gat ggg gac cca gcc tcc aac act gcc ccg ctc cag cca gag   1262
Ser Phe Asp Gly Asp Pro Ala Ser Asn Thr Ala Pro Leu Gln Pro Glu
380                 385                 390                 395

cag ctc caa gtg ttt gag act ctg gaa gag atc aca ggt tac cta tac   1310
Gln Leu Gln Val Phe Glu Thr Leu Glu Glu Ile Thr Gly Tyr Leu Tyr
```

```
                400                 405                 410

atc tca gca tgg ccg gac agc ctg cct gac ctc agc gtc ttc cag aac   1358
Ile Ser Ala Trp Pro Asp Ser Leu Pro Asp Leu Ser Val Phe Gln Asn
            415                 420                 425

ctg caa gta atc cgg gga cga att ctg cac aat ggc gcc tac tcg ctg   1406
Leu Gln Val Ile Arg Gly Arg Ile Leu His Asn Gly Ala Tyr Ser Leu
        430                 435                 440

acc ctg caa ggg ctg ggc atc agc tgg ctg ggg ctg cgc tca ctg agg   1454
Thr Leu Gln Gly Leu Gly Ile Ser Trp Leu Gly Leu Arg Ser Leu Arg
    445                 450                 455

gaa ctg ggc agt gga ctg gcc ctc atc cac cat aac acc cac ctc tgc   1502
Glu Leu Gly Ser Gly Leu Ala Leu Ile His His Asn Thr His Leu Cys
460                 465                 470                 475

ttc gtg cac acg gtg ccc tgg gac cag ctc ttt cgg aac ccg cac caa   1550
Phe Val His Thr Val Pro Trp Asp Gln Leu Phe Arg Asn Pro His Gln
                480                 485                 490

gct ctg ctc cac act gcc aac cgg cca gag gac gag tgt gtg ggc gag   1598
Ala Leu Leu His Thr Ala Asn Arg Pro Glu Asp Glu Cys Val Gly Glu
            495                 500                 505

ggc ctg gcc tgc cac cag ctg tgc gcc cga ggg cac tgc tgg ggt cca   1646
Gly Leu Ala Cys His Gln Leu Cys Ala Arg Gly His Cys Trp Gly Pro
        510                 515                 520

ggg ccc acc cag tgt gtc aac tgc agc cag ttc ctt cgg ggc cag gag   1694
Gly Pro Thr Gln Cys Val Asn Cys Ser Gln Phe Leu Arg Gly Gln Glu
    525                 530                 535

tgc gtg gag gaa tgc cga gta ctg cag ggg ctc ccc agg gag tat gtg   1742
Cys Val Glu Glu Cys Arg Val Leu Gln Gly Leu Pro Arg Glu Tyr Val
540                 545                 550                 555

aat gcc agg cac tgt ttg ccg tgc cac cct gag tgt cag ccc cag aat   1790
Asn Ala Arg His Cys Leu Pro Cys His Pro Glu Cys Gln Pro Gln Asn
                560                 565                 570

ggc tca gtg acc tgt ttt gga ccg gag gct gac cag tgt gtg gcc tgt   1838
Gly Ser Val Thr Cys Phe Gly Pro Glu Ala Asp Gln Cys Val Ala Cys
            575                 580                 585

gcc cac tat aag gac cct ccc ttc tgc gtg gcc cgc tgc ccc agc ggt   1886
Ala His Tyr Lys Asp Pro Pro Phe Cys Val Ala Arg Cys Pro Ser Gly
        590                 595                 600

gtg aaa cct gac ctc tcc tac atg ccc atc tgg aag ttt cca gat gag   1934
Val Lys Pro Asp Leu Ser Tyr Met Pro Ile Trp Lys Phe Pro Asp Glu
    605                 610                 615

gag ggc gca tgc cag cct tgc ccc atc aac tgc acc cac tcc cct ctg   1982
Glu Gly Ala Cys Gln Pro Cys Pro Ile Asn Cys Thr His Ser Pro Leu
620                 625                 630                 635

acg tcc atc gtc tct gcg gtg gtt ggc att ctg ctg gtc gtg gtc ttg   2030
Thr Ser Ile Val Ser Ala Val Val Gly Ile Leu Leu Val Val Val Leu
                640                 645                 650
```

```
ggg gtg gtc ttt ggg atc ctc atc aag cga cgg cag caa tcg aat tcc   2078
Gly Val Val Phe Gly Ile Leu Ile Lys Arg Arg Gln Gln Ser Asn Ser
            655                 660                 665

cgc ggc cgc cat ggc ggc cgg gag cat gcg acg tcg ggc cca att cgc   2126
Arg Gly Arg His Gly Gly Arg Glu His Ala Thr Ser Gly Pro Ile Arg
        670                 675                 680

cct ata                                                           2132
Pro Ile
    685
```

<210> 4
<211> 685
<212> PRT
<213> Homo sapiens

<400> 4

```
Met Glu Leu Ala Ala Leu Cys Arg Trp Gly Leu Leu Leu Ala Leu Leu
  1               5                  10                  15

Pro Pro Gly Ala Ala Ser Thr Gln Val Cys Thr Gly Thr Asp Met Lys
            20                  25                  30

Leu Arg Leu Pro Ala Ser Pro Glu Thr His Leu Asp Met Leu Arg His
        35                  40                  45

Leu Tyr Gln Gly Cys Gln Val Val Gln Gly Asn Leu Glu Leu Thr Tyr
    50                  55                  60

Leu Pro Thr Asn Ala Ser Leu Ser Phe Leu Gln Asp Ile Gln Glu Val
65                  70                  75                  80

Gln Gly Tyr Val Leu Ile Ala His Asn Gln Val Arg Gln Val Pro Leu
                85                  90                  95

Gln Arg Leu Arg Ile Val Arg Gly Thr Gln Leu Phe Glu Asp Asn Tyr
            100                 105                 110

Ala Leu Ala Val Leu Asp Asn Gly Asp Pro Leu Asn Asn Thr Thr Pro
        115                 120                 125

Val Thr Gly Ala Ser Pro Gly Gly Leu Arg Glu Leu Gln Leu Arg Ser
    130                 135                 140

Leu Thr Glu Ile Leu Lys Gly Gly Val Leu Ile Gln Arg Asn Pro Gln
145                 150                 155                 160

Leu Cys Tyr Gln Asp Thr Ile Leu Trp Lys Asp Ile Phe His Lys Asn
                165                 170                 175

Asn Gln Leu Ala Leu Thr Leu Ile Asp Thr Asn Arg Ser Arg Ala Cys
            180                 185                 190

His Pro Cys Ser Pro Met Cys Lys Gly Ser Arg Cys Trp Gly Glu Ser
        195                 200                 205

Ser Glu Asp Cys Gln Ser Leu Thr Arg Thr Val Cys Ala Gly Gly Cys
    210                 215                 220
```

Ala Arg Cys Lys Gly Pro Leu Pro Thr Asp Cys Cys His Glu Gln Cys
225 230 235 240

Ala Ala Gly Cys Thr Gly Pro Lys His Ser Asp Cys Leu Ala Cys Leu
245 250 255

His Phe Asn His Ser Gly Ile Cys Glu Leu His Cys Pro Ala Leu Val
260 265 270

Thr Tyr Asn Thr Asp Thr Phe Glu Ser Met Pro Asn Pro Glu Gly Arg
275 280 285

Tyr Thr Phe Gly Ala Ser Cys Val Thr Ala Cys Pro Tyr Asn Tyr Leu
290 295 300

Ser Thr Asp Val Gly Ser Cys Thr Leu Val Cys Pro Leu His Asn Gln
305 310 315 320

Glu Val Thr Ala Glu Asp Gly Thr Gln Arg Cys Glu Lys Cys Ser Lys
325 330 335

Pro Cys Ala Arg Val Cys Tyr Gly Leu Gly Met Glu His Leu Arg Glu
340 345 350

Val Arg Ala Val Thr Ser Ala Asn Ile Gln Glu Phe Ala Gly Cys Lys
355 360 365

Lys Ile Phe Gly Ser Leu Ala Phe Leu Pro Glu Ser Phe Asp Gly Asp
370 375 380

Pro Ala Ser Asn Thr Ala Pro Leu Gln Pro Glu Gln Leu Gln Val Phe
385 390 395 400

Glu Thr Leu Glu Glu Ile Thr Gly Tyr Leu Tyr Ile Ser Ala Trp Pro
405 410 415

Asp Ser Leu Pro Asp Leu Ser Val Phe Gln Asn Leu Gln Val Ile Arg
420 425 430

Gly Arg Ile Leu His Asn Gly Ala Tyr Ser Leu Thr Leu Gln Gly Leu
435 440 445

Gly Ile Ser Trp Leu Gly Leu Arg Ser Leu Arg Glu Leu Gly Ser Gly
450 455 460

Leu Ala Leu Ile His His Asn Thr His Leu Cys Phe Val His Thr Val
465 470 475 480

Pro Trp Asp Gln Leu Phe Arg Asn Pro His Gln Ala Leu Leu His Thr
485 490 495

Ala Asn Arg Pro Glu Asp Glu Cys Val Gly Glu Gly Leu Ala Cys His
500 505 510

Gln Leu Cys Ala Arg Gly His Cys Trp Gly Pro Gly Pro Thr Gln Cys
515 520 525

Val Asn Cys Ser Gln Phe Leu Arg Gly Gln Glu Cys Val Glu Glu Cys
530 535 540

```
Arg Val Leu Gln Gly Leu Pro Arg Glu Tyr Val Asn Ala Arg His Cys
545                 550                 555                 560

Leu Pro Cys His Pro Glu Cys Gln Pro Gln Asn Gly Ser Val Thr Cys
                565                 570                 575

Phe Gly Pro Glu Ala Asp Gln Cys Val Ala Cys Ala His Tyr Lys Asp
            580                 585                 590

Pro Pro Phe Cys Val Ala Arg Cys Pro Ser Gly Val Lys Pro Asp Leu
        595                 600                 605

Ser Tyr Met Pro Ile Trp Lys Phe Pro Asp Glu Glu Gly Ala Cys Gln
    610                 615                 620

Pro Cys Pro Ile Asn Cys Thr His Ser Pro Leu Thr Ser Ile Val Ser
625                 630                 635                 640

Ala Val Val Gly Ile Leu Leu Val Val Val Leu Gly Val Val Phe Gly
                645                 650                 655

Ile Leu Ile Lys Arg Arg Gln Gln Ser Asn Ser Arg Gly Arg His Gly
            660                 665                 670

Gly Arg Glu His Ala Thr Ser Gly Pro Ile Arg Pro Ile
        675                 680                 685
```

<210> 5
<211> 2164
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(2160)

<400> 5

```
atg gag ctg gcg gcc ttg tgc cgc tgg ggg ctc ctc ctc gcc ctc ttg   48
Met Glu Leu Ala Ala Leu Cys Arg Trp Gly Leu Leu Leu Ala Leu Leu
  1               5                  10                  15

ccc ccc gga gcc gcg agc acc caa gtg tgc acc ggc aca gac atg aag   96
Pro Pro Gly Ala Ala Ser Thr Gln Val Cys Thr Gly Thr Asp Met Lys
            20                  25                  30

ctg cgg ctc cct gcc agt ccc gag acc cac ctg gac atg ctc cgc cac  144
Leu Arg Leu Pro Ala Ser Pro Glu Thr His Leu Asp Met Leu Arg His
        35                  40                  45

ctc tac cag ggc tgc cag gtg gtg cag gga aac ctg gaa ctc acc tac  192
Leu Tyr Gln Gly Cys Gln Val Val Gln Gly Asn Leu Glu Leu Thr Tyr
    50                  55                  60

ctg ccc acc aat gcc agc ctg tcc ttc ctg cag gat atc cag gag gtg  240
Leu Pro Thr Asn Ala Ser Leu Ser Phe Leu Gln Asp Ile Gln Glu Val
65                  70                  75                  80

cag ggc tac gtg ctc atc gct cac aac caa gtg agg cag gtc cca ctg  288
Gln Gly Tyr Val Leu Ile Ala His Asn Gln Val Arg Gln Val Pro Leu
                85                  90                  95
```

```
cag agg ctg cgg att gtg cga ggc acc cag ctc ttt gag gac aac tat   336
Gln Arg Leu Arg Ile Val Arg Gly Thr Gln Leu Phe Glu Asp Asn Tyr
            100                 105                 110

gcc ctg gcc gtg cta gac aat gga gac ccg ctg aac aat acc acc cct   384
Ala Leu Ala Val Leu Asp Asn Gly Asp Pro Leu Asn Asn Thr Thr Pro
        115                 120                 125

gtc aca ggg gcc tcc cca gga ggc ctg cgg gag ctg cag ctt cga agc   432
Val Thr Gly Ala Ser Pro Gly Gly Leu Arg Glu Leu Gln Leu Arg Ser
    130                 135                 140

ctc aca gag atc ttg aaa gga ggg gtc ttg atc cag cgg aac ccc cag   480
Leu Thr Glu Ile Leu Lys Gly Gly Val Leu Ile Gln Arg Asn Pro Gln
145                 150                 155                 160

ctc tgc tac cag gac acg att ttg tgg aag gac atc ttc cac aag aac   528
Leu Cys Tyr Gln Asp Thr Ile Leu Trp Lys Asp Ile Phe His Lys Asn
                165                 170                 175

aac cag ctg gct ctc aca ctg ata gac acc aac cgc tct cgg gcc tgc   576
Asn Gln Leu Ala Leu Thr Leu Ile Asp Thr Asn Arg Ser Arg Ala Cys
            180                 185                 190

cac ccc tgt tct ccg atg tgt aag ggc tcc cgc tgc tgg gga gag agt   624
His Pro Cys Ser Pro Met Cys Lys Gly Ser Arg Cys Trp Gly Glu Ser
        195                 200                 205

tct gag gat tgt cag agc ctg acg cgc act gtc tgt gcc ggt ggc tgt   672
Ser Glu Asp Cys Gln Ser Leu Thr Arg Thr Val Cys Ala Gly Gly Cys
    210                 215                 220

gcc cgc tgc aag ggg cca ctg ccc act gac tgc tgc cat gag cag tgt   720
Ala Arg Cys Lys Gly Pro Leu Pro Thr Asp Cys Cys His Glu Gln Cys
225                 230                 235                 240

gct gcc ggc tgc acg ggc ccc aag cac tct gac tgc ctg gcc tgc ctc   768
Ala Ala Gly Cys Thr Gly Pro Lys His Ser Asp Cys Leu Ala Cys Leu
                245                 250                 255

cac ttc aac cac agt ggc atc tgt gag ctg cac tgc cca gcc ctg gtc   816
His Phe Asn His Ser Gly Ile Cys Glu Leu His Cys Pro Ala Leu Val
            260                 265                 270

acc tac aac aca gac acg ttt gag tcc atg ccc aat ccc gag ggc cgg   864
Thr Tyr Asn Thr Asp Thr Phe Glu Ser Met Pro Asn Pro Glu Gly Arg
        275                 280                 285

tat aca ttc ggc gcc agc tgt gtg act gcc tgt ccc tac aac tac ctt   912
Tyr Thr Phe Gly Ala Ser Cys Val Thr Ala Cys Pro Tyr Asn Tyr Leu
    290                 295                 300

tct acg gac gtg gga tcc tgc acc ctc gtc tgc ccc ctg cac aac caa   960
Ser Thr Asp Val Gly Ser Cys Thr Leu Val Cys Pro Leu His Asn Gln
305                 310                 315                 320

gag gtg aca gca gag gat gga aca cag cgg tgt gag aag tgc agc aag  1008
Glu Val Thr Ala Glu Asp Gly Thr Gln Arg Cys Glu Lys Cys Ser Lys
                325                 330                 335
```

```
ccc tgt gcc cga gtg tgc tat ggt ctg ggc atg gag cac ttg cga gag   1056
Pro Cys Ala Arg Val Cys Tyr Gly Leu Gly Met Glu His Leu Arg Glu
            340                 345                 350

gtg agg gca gtt acc agt gcc aat atc cag gag ttt gct ggc tgc aag   1104
Val Arg Ala Val Thr Ser Ala Asn Ile Gln Glu Phe Ala Gly Cys Lys
        355                 360                 365

aag atc ttt ggg agc ctg gca ttt ctg ccg gag agc ttt gat gga gtc   1152
Lys Ile Phe Gly Ser Leu Ala Phe Leu Pro Glu Ser Phe Asp Gly Val
    370                 375                 380

tca ctc tgt cag cag gct gga gtg cag tgg tac gat ctt ggc tca ctg   1200
Ser Leu Cys Gln Gln Ala Gly Val Gln Trp Tyr Asp Leu Gly Ser Leu
385                 390                 395                 400

caa cct ctg cct cct gga ttc aag caa ttc tcc tgc ctc agt ctc ctg   1248
Gln Pro Leu Pro Pro Gly Phe Lys Gln Phe Ser Cys Leu Ser Leu Leu
                405                 410                 415

agt agc tgg gac tac agg gac cca gcc tcc aac act gcc ccg ctc cag   1296
Ser Ser Trp Asp Tyr Arg Asp Pro Ala Ser Asn Thr Ala Pro Leu Gln
            420                 425                 430

cca gag cag ctc caa gtg ttt gag act ctg gaa gag atc aca ggt tac   1344
Pro Glu Gln Leu Gln Val Phe Glu Thr Leu Glu Glu Ile Thr Gly Tyr
        435                 440                 445

cta tac atc tca gca tgg ccg gac agc ctg cct gac ctc agc gtc ttc   1392
Leu Tyr Ile Ser Ala Trp Pro Asp Ser Leu Pro Asp Leu Ser Val Phe
    450                 455                 460

cag aac ctg caa gta atc cgg gga cga att ctg cac aat ggc gcc tac   1440
Gln Asn Leu Gln Val Ile Arg Gly Arg Ile Leu His Asn Gly Ala Tyr
465                 470                 475                 480

tcg ctg acc ctg caa ggg ctg ggc atc agc tgg ctg ggg ctg cgc tca   1488
Ser Leu Thr Leu Gln Gly Leu Gly Ile Ser Trp Leu Gly Leu Arg Ser
                485                 490                 495

ctg agg gaa ctg ggc agt gga ctg gcc ctc atc cac cat aac acc cac   1536
Leu Arg Glu Leu Gly Ser Gly Leu Ala Leu Ile His His Asn Thr His
            500                 505                 510

ctc tgc ttc gtg cac acg gtg ccc tgg gac cag ctc ttt cgg aac ccg   1584
Leu Cys Phe Val His Thr Val Pro Trp Asp Gln Leu Phe Arg Asn Pro
        515                 520                 525

cac caa gct ctg ctc cac act gcc aac cgg cca gag gac gag tgt gtg   1632
His Gln Ala Leu Leu His Thr Ala Asn Arg Pro Glu Asp Glu Cys Val
    530                 535                 540

ggc gag ggc ctg gcc tgc cac cag ctg tgc gcc cga ggg cac tgc tgg   1680
Gly Glu Gly Leu Ala Cys His Gln Leu Cys Ala Arg Gly His Cys Trp
545                 550                 555                 560

ggt cca ggg ccc acc cag tgt gtc aac tgc agc cag ttc ctt cgg ggc   1728
Gly Pro Gly Pro Thr Gln Cys Val Asn Cys Ser Gln Phe Leu Arg Gly
                565                 570                 575

cag gag tgc gtg gag gaa tgc cga gta ctg cag ggg ctc ccc agg gag   1776
```

```
Gln Glu Cys Val Glu Glu Cys Arg Val Leu Gln Gly Leu Pro Arg Glu
            580                 585                 590

tat gtg aat gcc agg cac tgt ttg ccg tgc cac cct gag tgt cag ccc   1824
Tyr Val Asn Ala Arg His Cys Leu Pro Cys His Pro Glu Cys Gln Pro
        595                 600                 605

cag aat ggc tca gtg acc tgt ttt gga ccg gag gct gac cag tgt gtg   1872
Gln Asn Gly Ser Val Thr Cys Phe Gly Pro Glu Ala Asp Gln Cys Val
    610                 615                 620

gcc tgt gcc cac tat aag gac cct ccc ttc tgc gtg gcc cgc tgc ccc   1920
Ala Cys Ala His Tyr Lys Asp Pro Pro Phe Cys Val Ala Arg Cys Pro
625                 630                 635                 640

agc ggt gtg aaa cct gac ctc tcc tac atg ccc atc tgg aag ttt cca   1968
Ser Gly Val Lys Pro Asp Leu Ser Tyr Met Pro Ile Trp Lys Phe Pro
                645                 650                 655

gat gag gag ggc gca tgc cag cct tgc ccc atc aac tgc acc cac tcc   2016
Asp Glu Glu Gly Ala Cys Gln Pro Cys Pro Ile Asn Cys Thr His Ser
            660                 665                 670

tgt gtg gac ctg gat gac aag ggc tgc ccc gcc gag cag aga gcc agc   2064
Cys Val Asp Leu Asp Asp Lys Gly Cys Pro Ala Glu Gln Arg Ala Ser
        675                 680                 685

cct ctg acg tcc atc atc tct gcg gtg gtt ggc att ctg ctg gtc gtg   2112
Pro Leu Thr Ser Ile Ile Ser Ala Val Val Gly Ile Leu Leu Val Val
    690                 695                 700

gtc ttg ggg gtg gtc ttt ggg atc ctc atc agc gac ggc agc aat cac   2160
Val Leu Gly Val Val Phe Gly Ile Leu Ile Ser Asp Gly Ser Asn His
705                 710                 715                 720

tagt                                                              2164
```

<210> 6
<211> 720
<212> PRT
<213> Homo sapiens

<400> 6

```
Met Glu Leu Ala Ala Leu Cys Arg Trp Gly Leu Leu Leu Ala Leu Leu
  1               5                  10                  15

Pro Pro Gly Ala Ala Ser Thr Gln Val Cys Thr Gly Thr Asp Met Lys
            20                  25                  30

Leu Arg Leu Pro Ala Ser Pro Glu Thr His Leu Asp Met Leu Arg His
        35                  40                  45

Leu Tyr Gln Gly Cys Gln Val Val Gln Gly Asn Leu Glu Leu Thr Tyr
    50                  55                  60

Leu Pro Thr Asn Ala Ser Leu Ser Phe Leu Gln Asp Ile Gln Glu Val
65                  70                  75                  80

Gln Gly Tyr Val Leu Ile Ala His Asn Gln Val Arg Gln Val Pro Leu
                85                  90                  95
```

Gln Arg Leu Arg Ile Val Arg Gly Thr Gln Leu Phe Glu Asp Asn Tyr
100 105 110

Ala Leu Ala Val Leu Asp Asn Gly Asp Pro Leu Asn Asn Thr Thr Pro
115 120 125

Val Thr Gly Ala Ser Pro Gly Gly Leu Arg Glu Leu Gln Leu Arg Ser
130 135 140

Leu Thr Glu Ile Leu Lys Gly Gly Val Leu Ile Gln Arg Asn Pro Gln
145 150 155 160

Leu Cys Tyr Gln Asp Thr Ile Leu Trp Lys Asp Ile Phe His Lys Asn
165 170 175

Asn Gln Leu Ala Leu Thr Leu Ile Asp Thr Asn Arg Ser Arg Ala Cys
180 185 190

His Pro Cys Ser Pro Met Cys Lys Gly Ser Arg Cys Trp Gly Glu Ser
195 200 205

Ser Glu Asp Cys Gln Ser Leu Thr Arg Thr Val Cys Ala Gly Gly Cys
210 215 220

Ala Arg Cys Lys Gly Pro Leu Pro Thr Asp Cys Cys His Glu Gln Cys
225 230 235 240

Ala Ala Gly Cys Thr Gly Pro Lys His Ser Asp Cys Leu Ala Cys Leu
245 250 255

His Phe Asn His Ser Gly Ile Cys Glu Leu His Cys Pro Ala Leu Val
260 265 270

Thr Tyr Asn Thr Asp Thr Phe Glu Ser Met Pro Asn Pro Glu Gly Arg
275 280 285

Tyr Thr Phe Gly Ala Ser Cys Val Thr Ala Cys Pro Tyr Asn Tyr Leu
290 295 300

Ser Thr Asp Val Gly Ser Cys Thr Leu Val Cys Pro Leu His Asn Gln
305 310 315 320

Glu Val Thr Ala Glu Asp Gly Thr Gln Arg Cys Glu Lys Cys Ser Lys
325 330 335

Pro Cys Ala Arg Val Cys Tyr Gly Leu Gly Met Glu His Leu Arg Glu
340 345 350

Val Arg Ala Val Thr Ser Ala Asn Ile Gln Glu Phe Ala Gly Cys Lys
355 360 365

Lys Ile Phe Gly Ser Leu Ala Phe Leu Pro Glu Ser Phe Asp Gly Val
370 375 380

Ser Leu Cys Gln Gln Ala Gly Val Gln Trp Tyr Asp Leu Gly Ser Leu
385 390 395 400

Gln Pro Leu Pro Pro Gly Phe Lys Gln Phe Ser Cys Leu Ser Leu Leu
405 410 415

```
Ser Ser Trp Asp Tyr Arg Asp Pro Ala Ser Asn Thr Ala Pro Leu Gln
            420                 425                 430

Pro Glu Gln Leu Gln Val Phe Glu Thr Leu Glu Glu Ile Thr Gly Tyr
        435                 440                 445

Leu Tyr Ile Ser Ala Trp Pro Asp Ser Leu Pro Asp Leu Ser Val Phe
    450                 455                 460

Gln Asn Leu Gln Val Ile Arg Gly Arg Ile Leu His Asn Gly Ala Tyr
465                 470                 475                 480

Ser Leu Thr Leu Gln Gly Leu Gly Ile Ser Trp Leu Gly Leu Arg Ser
                485                 490                 495

Leu Arg Glu Leu Gly Ser Gly Leu Ala Leu Ile His His Asn Thr His
            500                 505                 510

Leu Cys Phe Val His Thr Val Pro Trp Asp Gln Leu Phe Arg Asn Pro
        515                 520                 525

His Gln Ala Leu Leu His Thr Ala Asn Arg Pro Glu Asp Glu Cys Val
    530                 535                 540

Gly Glu Gly Leu Ala Cys His Gln Leu Cys Ala Arg Gly His Cys Trp
545                 550                 555                 560

Gly Pro Gly Pro Thr Gln Cys Val Asn Cys Ser Gln Phe Leu Arg Gly
                565                 570                 575

Gln Glu Cys Val Glu Glu Cys Arg Val Leu Gln Gly Leu Pro Arg Glu
            580                 585                 590

Tyr Val Asn Ala Arg His Cys Leu Pro Cys His Pro Glu Cys Gln Pro
        595                 600                 605

Gln Asn Gly Ser Val Thr Cys Phe Gly Pro Glu Ala Asp Gln Cys Val
    610                 615                 620

Ala Cys Ala His Tyr Lys Asp Pro Pro Phe Cys Val Ala Arg Cys Pro
625                 630                 635                 640

Ser Gly Val Lys Pro Asp Leu Ser Tyr Met Pro Ile Trp Lys Phe Pro
                645                 650                 655

Asp Glu Glu Gly Ala Cys Gln Pro Cys Pro Ile Asn Cys Thr His Ser
            660                 665                 670

Cys Val Asp Leu Asp Asp Lys Gly Cys Pro Ala Glu Gln Arg Ala Ser
        675                 680                 685

Pro Leu Thr Ser Ile Ile Ser Ala Val Val Gly Ile Leu Leu Val Val
    690                 695                 700

Val Leu Gly Val Val Phe Gly Ile Leu Ile Ser Asp Gly Ser Asn His
705                 710                 715                 720
```

<210> 7
<211> 884
<212> DNA

<213> Homo sapiens

<220>
<221> CDS
<222> (77)..(568)

<400> 7
acgcgttggg agctctccat atggtcgacc tgcaggcggc cgcgaattca ctagtgattg 60

agccgcagtg agcacc atg gag ctg gcg gcc ttg tgc cgc tgg ggg ctc ctc 112
Met Glu Leu Ala Ala Leu Cys Arg Trp Gly Leu Leu
1 5 10

ctc gcc ctc ttg ccc ccc gga gcc gcg agc acc caa gtg tgc acc ggc 160
Leu Ala Leu Leu Pro Pro Gly Ala Ala Ser Thr Gln Val Cys Thr Gly
15 20 25

aca gac atg aag ctg cgg ctc cct gcc agt ccc gag acc cac ctg gac 208
Thr Asp Met Lys Leu Arg Leu Pro Ala Ser Pro Glu Thr His Leu Asp
30 35 40

atg ctc cgc cac ctc tac cag ggc tgc cag gtg gtg cag gga aac ctg 256
Met Leu Arg His Leu Tyr Gln Gly Cys Gln Val Val Gln Gly Asn Leu
45 50 55 60

gaa ctc acc tac ctg ccc acc aat gcc agc ctg tcc ttc ctg cag gat 304
Glu Leu Thr Tyr Leu Pro Thr Asn Ala Ser Leu Ser Phe Leu Gln Asp
65 70 75

atc cag gag gtg cag ggc tac gtg ctc atc gct cac aac caa gtg agg 352
Ile Gln Glu Val Gln Gly Tyr Val Leu Ile Ala His Asn Gln Val Arg
80 85 90

cag gtc cca ctg cag agg ctg cgg att gtg cga ggc acc cag ctc ttt 400
Gln Val Pro Leu Gln Arg Leu Arg Ile Val Arg Gly Thr Gln Leu Phe
95 100 105

gag gac aac tat gcc ctg gcc gtg cta gac aat gga gac ccg ctg aac 448
Glu Asp Asn Tyr Ala Leu Ala Val Leu Asp Asn Gly Asp Pro Leu Asn
110 115 120

aat acc acc cct gtc aca ggg gcc tcc cca gga ggc ctg cgg gag ctg 496
Asn Thr Thr Pro Val Thr Gly Ala Ser Pro Gly Gly Leu Arg Glu Leu
125 130 135 140

cag ctt cga agc ctc aca gag atc ttg aaa gga ggg gtc ttg atc cag 544
Gln Leu Arg Ser Leu Thr Glu Ile Leu Lys Gly Gly Val Leu Ile Gln
145 150 155

cgg aac ccc cag cgg tgt gaa acc tgacctctcc tacatgccca tctggaagtt 598
Arg Asn Pro Gln Arg Cys Glu Thr
160

tccagatgag gagggcgcat gccagccttg ccccatcaac tgcacccact cctgtgtgga 658

cctggatgac aagggctgcc ccgccgagca gagagccagc cctctgacgt ccatcatctc 718

tgcggtggtt ggcattctgc tggtcgtggt cttgggggtg gtctttggga tcctcatcaa 778

gcgacggcag caatcgaatt cccgcggccg ccatggcggc cgggagcatg cgacgtcggg 838

cccaattcgc cctatagtga gtcgtattac aattcactgg ccgtcg 884

<210> 8
<211> 164
<212> PRT
<213> Homo sapiens

<400> 8
Met Glu Leu Ala Ala Leu Cys Arg Trp Gly Leu Leu Leu Ala Leu Leu
1 5 10 15

Pro Pro Gly Ala Ala Ser Thr Gln Val Cys Thr Gly Thr Asp Met Lys
20 25 30

Leu Arg Leu Pro Ala Ser Pro Glu Thr His Leu Asp Met Leu Arg His
35 40 45

Leu Tyr Gln Gly Cys Gln Val Val Gln Gly Asn Leu Glu Leu Thr Tyr
50 55 60

Leu Pro Thr Asn Ala Ser Leu Ser Phe Leu Gln Asp Ile Gln Glu Val
65 70 75 80

Gln Gly Tyr Val Leu Ile Ala His Asn Gln Val Arg Gln Val Pro Leu
85 90 95

Gln Arg Leu Arg Ile Val Arg Gly Thr Gln Leu Phe Glu Asp Asn Tyr
100 105 110

Ala Leu Ala Val Leu Asp Asn Gly Asp Pro Leu Asn Asn Thr Thr Pro
115 120 125

Val Thr Gly Ala Ser Pro Gly Gly Leu Arg Glu Leu Gln Leu Arg Ser
130 135 140

Leu Thr Glu Ile Leu Lys Gly Gly Val Leu Ile Gln Arg Asn Pro Gln
145 150 155 160

Arg Cys Glu Thr

<210> 9
<211> 2149
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(2145)

<400> 9
atg gag ctg gcg gcc ttg tgc cgc tgg ggg ctc ctc ctc gcc ctc ttg 48
Met Glu Leu Ala Ala Leu Cys Arg Trp Gly Leu Leu Leu Ala Leu Leu
1 5 10 15

ccc ccc gga gcc gcg agc acc caa gtg tgc acc ggc aca gac atg aag 96
Pro Pro Gly Ala Ala Ser Thr Gln Val Cys Thr Gly Thr Asp Met Lys
20 25 30

```
ctg cgg ctc cct gcc agt ccc gag acc cac ctg gac atg ctc cgc cac   144
Leu Arg Leu Pro Ala Ser Pro Glu Thr His Leu Asp Met Leu Arg His
         35                  40                  45

ctc tac cag ggc tgc cag gtg gtg cag gga aac ctg gaa ctc acc tac   192
Leu Tyr Gln Gly Cys Gln Val Val Gln Gly Asn Leu Glu Leu Thr Tyr
     50                  55                  60

ctg ccc acc aat gcc agc ctg tcc ttc ctg cag gat atc cag gag gtg   240
Leu Pro Thr Asn Ala Ser Leu Ser Phe Leu Gln Asp Ile Gln Glu Val
 65                  70                  75                  80

cag ggc tac gtg ctc atc gct cac aac caa gtg agg cag gtc cca ctg   288
Gln Gly Tyr Val Leu Ile Ala His Asn Gln Val Arg Gln Val Pro Leu
                 85                  90                  95

cag agg ctg cgg att gtg cga ggc acc cag ctc ttt gag gac aac tat   336
Gln Arg Leu Arg Ile Val Arg Gly Thr Gln Leu Phe Glu Asp Asn Tyr
            100                 105                 110

gcc ctg gcc gtg cta gac aat gga gac ccg ctg aac aat acc acc cct   384
Ala Leu Ala Val Leu Asp Asn Gly Asp Pro Leu Asn Asn Thr Thr Pro
        115                 120                 125

gtc aca ggg gcc tcc cca gga ggc ctg cgg gag ctg cag ctt cga agc   432
Val Thr Gly Ala Ser Pro Gly Gly Leu Arg Glu Leu Gln Leu Arg Ser
    130                 135                 140

ctc aca gag atc ttg aaa gga ggg gtc ttg atc cag cgg aac ccc cag   480
Leu Thr Glu Ile Leu Lys Gly Gly Val Leu Ile Gln Arg Asn Pro Gln
145                 150                 155                 160

ctc tgc tac cag gac acg att ttg tgg aag gac atc ttc cac aag aac   528
Leu Cys Tyr Gln Asp Thr Ile Leu Trp Lys Asp Ile Phe His Lys Asn
                165                 170                 175

aac cag ctg gct ctc aca ctg ata gac acc aac cgc tct cgg gcc tgc   576
Asn Gln Leu Ala Leu Thr Leu Ile Asp Thr Asn Arg Ser Arg Ala Cys
            180                 185                 190

cac ccc tgt tct ccg atg tgt aag ggc tcc cgc tgc tgg gga gag agt   624
His Pro Cys Ser Pro Met Cys Lys Gly Ser Arg Cys Trp Gly Glu Ser
        195                 200                 205

tct gag gat tgt cag agc ctg acg cgc act gtc tgt gcc ggt ggc tgt   672
Ser Glu Asp Cys Gln Ser Leu Thr Arg Thr Val Cys Ala Gly Gly Cys
    210                 215                 220

gcc cgc tgc aag ggg cca ctg ccc act gac tgc tgc cat gag cag tgt   720
Ala Arg Cys Lys Gly Pro Leu Pro Thr Asp Cys Cys His Glu Gln Cys
225                 230                 235                 240

gct gcc ggc tgc acg ggc ccc aag cac tct gac tgc ctg gcc tgc ctc   768
Ala Ala Gly Cys Thr Gly Pro Lys His Ser Asp Cys Leu Ala Cys Leu
                245                 250                 255

cac ttc aac cac agt ggc atc tgt gag ctg cac tgc cca gcc ctg gtc   816
His Phe Asn His Ser Gly Ile Cys Glu Leu His Cys Pro Ala Leu Val
            260                 265                 270

acc tac aac aca gac acg ttt gag tcc atg ccc aat ccc gag ggc cgg   864
```

```
Thr Tyr Asn Thr Asp Thr Phe Glu Ser Met Pro Asn Pro Glu Gly Arg
        275                 280                 285

tat aca ttc ggc gcc agc tgt gtg act gcc tgt ccc tac aac tac ctt   912
Tyr Thr Phe Gly Ala Ser Cys Val Thr Ala Cys Pro Tyr Asn Tyr Leu
    290                 295                 300

tct acg gac gtg gga tcc tgc acc ctc gtc tgc ccc ctg cac aac caa   960
Ser Thr Asp Val Gly Ser Cys Thr Leu Val Cys Pro Leu His Asn Gln
305                 310                 315                 320

gag gtg aca gca gag gat gga aca cag cgg tgt gag aag tgc agc aag   1008
Glu Val Thr Ala Glu Asp Gly Thr Gln Arg Cys Glu Lys Cys Ser Lys
                325                 330                 335

ccc tgt gcc cga gtg tgc tat ggt ctg ggc atg gag cac ttg cga gag   1056
Pro Cys Ala Arg Val Cys Tyr Gly Leu Gly Met Glu His Leu Arg Glu
            340                 345                 350

gtg agg gca gtt acc agt gcc aat atc cag gag ttt gct ggc tgc aag   1104
Val Arg Ala Val Thr Ser Ala Asn Ile Gln Glu Phe Ala Gly Cys Lys
        355                 360                 365

aag atc ttt ggg agc ctg gca ttt ctg ccg gag agc ttt gat ggg gac   1152
Lys Ile Phe Gly Ser Leu Ala Phe Leu Pro Glu Ser Phe Asp Gly Asp
    370                 375                 380

cca gcc tcc aac act gcc ccg ctc cag cca gag cag ctc caa gtg ttt   1200
Pro Ala Ser Asn Thr Ala Pro Leu Gln Pro Glu Gln Leu Gln Val Phe
385                 390                 395                 400

gag act ctg gaa gag atc aca ggt tac cta tac atc tca gca tgg ccg   1248
Glu Thr Leu Glu Glu Ile Thr Gly Tyr Leu Tyr Ile Ser Ala Trp Pro
                405                 410                 415

gac agc ctg cct gac ctc agc gtc ttc cag aac ctg caa gta atc cgg   1296
Asp Ser Leu Pro Asp Leu Ser Val Phe Gln Asn Leu Gln Val Ile Arg
            420                 425                 430

gga cga att ctg cac aat ggc gcc tac tcg ctg acc ctg caa ggg ctg   1344
Gly Arg Ile Leu His Asn Gly Ala Tyr Ser Leu Thr Leu Gln Gly Leu
        435                 440                 445

ggc atc agc tgg ctg ggg ctg cgc tca ctg agg gaa ctg ggc agt gga   1392
Gly Ile Ser Trp Leu Gly Leu Arg Ser Leu Arg Glu Leu Gly Ser Gly
    450                 455                 460

ctg gcc ctc atc cac cat aac acc cac ctc tgc ttc gtg cac acg gtg   1440
Leu Ala Leu Ile His His Asn Thr His Leu Cys Phe Val His Thr Val
465                 470                 475                 480

ccc tgg gac cag ctc ttt cgg aac ccg cac caa gct ctg ctc cac act   1488
Pro Trp Asp Gln Leu Phe Arg Asn Pro His Gln Ala Leu Leu His Thr
                485                 490                 495

gcc aac cgg cca gag gac gag tgt gtg ggc gag ggc ctg gcc tgc cac   1536
Ala Asn Arg Pro Glu Asp Glu Cys Val Gly Glu Gly Leu Ala Cys His
            500                 505                 510

cag ctg tgc gcc cga ggg cac tgc tgg ggt cca ggg ccc acc cag tgt   1584
Gln Leu Cys Ala Arg Gly His Cys Trp Gly Pro Gly Pro Thr Gln Cys
```

```
        515                 520                 525

gtc aac tgc agc cag ttc ctt cgg ggc cag gag tgc gtg gag gaa tgc   1632
Val Asn Cys Ser Gln Phe Leu Arg Gly Gln Glu Cys Val Glu Glu Cys
    530                 535                 540

cga gta ctg cag ggg ctc ccc agg gag tat gtg aat gcc agg cac tgt   1680
Arg Val Leu Gln Gly Leu Pro Arg Glu Tyr Val Asn Ala Arg His Cys
545                 550                 555                 560

ttg ccg tgc cac cct gag tgt cag ccc cag aat ggc tca gtg acc tgt   1728
Leu Pro Cys His Pro Glu Cys Gln Pro Gln Asn Gly Ser Val Thr Cys
                565                 570                 575

ttt gga ccg gta atg cgt ttt cct ctc tgg gtg cct ccc att ttc tgg   1776
Phe Gly Pro Val Met Arg Phe Pro Leu Trp Val Pro Pro Ile Phe Trp
            580                 585                 590

ctc aag tcc ctg ccc agg atc aag ctt gga gga ggg ccc cga ggg agg   1824
Leu Lys Ser Leu Pro Arg Ile Lys Leu Gly Gly Gly Pro Arg Gly Arg
        595                 600                 605

ggc cac aga gac tgg gag gct gac cag tgt gtg gcc tgt gcc cac tat   1872
Gly His Arg Asp Trp Glu Ala Asp Gln Cys Val Ala Cys Ala His Tyr
    610                 615                 620

aag gac cct ccc ttc tgc gtg gcc cga tgc ccc agc ggt gtg aaa cct   1920
Lys Asp Pro Pro Phe Cys Val Ala Arg Cys Pro Ser Gly Val Lys Pro
625                 630                 635                 640

gac ctc tcc tac atg ccc atc tgg aag ttt cca gat gag gag ggc gca   1968
Asp Leu Ser Tyr Met Pro Ile Trp Lys Phe Pro Asp Glu Glu Gly Ala
                645                 650                 655

tgc cag cct tgc ccc atc aac tgc acc cac tcc tgt gtg gac ctg gat   2016
Cys Gln Pro Cys Pro Ile Asn Cys Thr His Ser Cys Val Asp Leu Asp
            660                 665                 670

gac aag ggc tgc ccc gcc gag cag aga gcc agc cct ctg atg tcc atc   2064
Asp Lys Gly Cys Pro Ala Glu Gln Arg Ala Ser Pro Leu Met Ser Ile
        675                 680                 685

atc tct gcg gtg gtt ggc att ctg ctg gtc gtg gtc ttg ggg gtg gtc   2112
Ile Ser Ala Val Val Gly Ile Leu Leu Val Val Val Leu Gly Val Val
    690                 695                 700

ttt ggg atc ctc ata agc gac ggc agc aat cac tagt                  2149
Phe Gly Ile Leu Ile Ser Asp Gly Ser Asn His
705                 710                 715
```

<210> 10
<211> 715
<212> PRT
<213> Homo sapiens

<400> 10

```
Met Glu Leu Ala Ala Leu Cys Arg Trp Gly Leu Leu Leu Ala Leu Leu
  1               5                  10                  15

Pro Pro Gly Ala Ala Ser Thr Gln Val Cys Thr Gly Thr Asp Met Lys
```

```
          20                  25                  30

Leu Arg Leu Pro Ala Ser Pro Glu Thr His Leu Asp Met Leu Arg His
        35                  40                  45

Leu Tyr Gln Gly Cys Gln Val Val Gln Gly Asn Leu Glu Leu Thr Tyr
    50                  55                  60

Leu Pro Thr Asn Ala Ser Leu Ser Phe Leu Gln Asp Ile Gln Glu Val
65                  70                  75                  80

Gln Gly Tyr Val Leu Ile Ala His Asn Gln Val Arg Gln Val Pro Leu
                85                  90                  95

Gln Arg Leu Arg Ile Val Arg Gly Thr Gln Leu Phe Glu Asp Asn Tyr
            100                 105                 110

Ala Leu Ala Val Leu Asp Asn Gly Asp Pro Leu Asn Asn Thr Thr Pro
        115                 120                 125

Val Thr Gly Ala Ser Pro Gly Gly Leu Arg Glu Leu Gln Leu Arg Ser
    130                 135                 140

Leu Thr Glu Ile Leu Lys Gly Gly Val Leu Ile Gln Arg Asn Pro Gln
145                 150                 155                 160

Leu Cys Tyr Gln Asp Thr Ile Leu Trp Lys Asp Ile Phe His Lys Asn
                165                 170                 175

Asn Gln Leu Ala Leu Thr Leu Ile Asp Thr Asn Arg Ser Arg Ala Cys
            180                 185                 190

His Pro Cys Ser Pro Met Cys Lys Gly Ser Arg Cys Trp Gly Glu Ser
        195                 200                 205

Ser Glu Asp Cys Gln Ser Leu Thr Arg Thr Val Cys Ala Gly Gly Cys
    210                 215                 220

Ala Arg Cys Lys Gly Pro Leu Pro Thr Asp Cys Cys His Glu Gln Cys
225                 230                 235                 240

Ala Ala Gly Cys Thr Gly Pro Lys His Ser Asp Cys Leu Ala Cys Leu
                245                 250                 255

His Phe Asn His Ser Gly Ile Cys Glu Leu His Cys Pro Ala Leu Val
            260                 265                 270

Thr Tyr Asn Thr Asp Thr Phe Glu Ser Met Pro Asn Pro Glu Gly Arg
        275                 280                 285

Tyr Thr Phe Gly Ala Ser Cys Val Thr Ala Cys Pro Tyr Asn Tyr Leu
    290                 295                 300

Ser Thr Asp Val Gly Ser Cys Thr Leu Val Cys Pro Leu His Asn Gln
305                 310                 315                 320

Glu Val Thr Ala Glu Asp Gly Thr Gln Arg Cys Glu Lys Cys Ser Lys
                325                 330                 335

Pro Cys Ala Arg Val Cys Tyr Gly Leu Gly Met Glu His Leu Arg Glu
            340                 345                 350
```

Val Arg Ala Val Thr Ser Ala Asn Ile Gln Glu Phe Ala Gly Cys Lys
355 360 365

Lys Ile Phe Gly Ser Leu Ala Phe Leu Pro Glu Ser Phe Asp Gly Asp
370 375 380

Pro Ala Ser Asn Thr Ala Pro Leu Gln Pro Glu Gln Leu Gln Val Phe
385 390 395 400

Glu Thr Leu Glu Glu Ile Thr Gly Tyr Leu Tyr Ile Ser Ala Trp Pro
405 410 415

Asp Ser Leu Pro Asp Leu Ser Val Phe Gln Asn Leu Gln Val Ile Arg
420 425 430

Gly Arg Ile Leu His Asn Gly Ala Tyr Ser Leu Thr Leu Gln Gly Leu
435 440 445

Gly Ile Ser Trp Leu Gly Leu Arg Ser Leu Arg Glu Leu Gly Ser Gly
450 455 460

Leu Ala Leu Ile His His Asn Thr His Leu Cys Phe Val His Thr Val
465 470 475 480

Pro Trp Asp Gln Leu Phe Arg Asn Pro His Gln Ala Leu Leu His Thr
485 490 495

Ala Asn Arg Pro Glu Asp Glu Cys Val Gly Glu Gly Leu Ala Cys His
500 505 510

Gln Leu Cys Ala Arg Gly His Cys Trp Gly Pro Gly Pro Thr Gln Cys
515 520 525

Val Asn Cys Ser Gln Phe Leu Arg Gly Gln Glu Cys Val Glu Glu Cys
530 535 540

Arg Val Leu Gln Gly Leu Pro Arg Glu Tyr Val Asn Ala Arg His Cys
545 550 555 560

Leu Pro Cys His Pro Glu Cys Gln Pro Gln Asn Gly Ser Val Thr Cys
565 570 575

Phe Gly Pro Val Met Arg Phe Pro Leu Trp Val Pro Pro Ile Phe Trp
580 585 590

Leu Lys Ser Leu Pro Arg Ile Lys Leu Gly Gly Gly Pro Arg Gly Arg
595 600 605

Gly His Arg Asp Trp Glu Ala Asp Gln Cys Val Ala Cys Ala His Tyr
610 615 620

Lys Asp Pro Pro Phe Cys Val Ala Arg Cys Pro Ser Gly Val Lys Pro
625 630 635 640

Asp Leu Ser Tyr Met Pro Ile Trp Lys Phe Pro Asp Glu Glu Gly Ala
645 650 655

Cys Gln Pro Cys Pro Ile Asn Cys Thr His Ser Cys Val Asp Leu Asp
660 665 670

Asp Lys Gly Cys Pro Ala Glu Gln Arg Ala Ser Pro Leu Met Ser Ile
675 680 685

Ile Ser Ala Val Val Gly Ile Leu Leu Val Val Val Leu Gly Val Val
690 695 700

Phe Gly Ile Leu Ile Ser Asp Gly Ser Asn His
705 710 715

<210> 11
<211> 690
<212> PRT
<213> Homo sapiens

<400> 11
Met Glu Leu Ala Ala Leu Cys Arg Trp Gly Leu Leu Leu Ala Leu Leu
1 5 10 15

Pro Pro Gly Ala Ala Ser Thr Gln Val Cys Thr Gly Thr Asp Met Lys
20 25 30

Leu Arg Leu Pro Ala Ser Pro Glu Thr His Leu Asp Met Leu Arg His
35 40 45

Leu Tyr Gln Gly Cys Gln Val Val Gln Gly Asn Leu Glu Leu Thr Tyr
50 55 60

Leu Pro Thr Asn Ala Ser Leu Ser Phe Leu Gln Asp Ile Gln Glu Val
65 70 75 80

Gln Gly Tyr Val Leu Ile Ala His Asn Gln Val Arg Gln Val Pro Leu
85 90 95

Gln Arg Leu Arg Ile Val Arg Gly Thr Gln Leu Phe Glu Asp Asn Tyr
100 105 110

Ala Leu Ala Val Leu Asp Asn Gly Asp Pro Leu Asn Asn Thr Thr Pro
115 120 125

Val Thr Gly Ala Ser Pro Gly Gly Leu Arg Glu Leu Gln Leu Arg Ser
130 135 140

Leu Thr Glu Ile Leu Lys Gly Gly Val Leu Ile Gln Arg Asn Pro Gln
145 150 155 160

Leu Cys Tyr Gln Asp Thr Ile Leu Trp Lys Asp Ile Phe His Lys Asn
165 170 175

Asn Gln Leu Ala Leu Thr Leu Ile Asp Thr Asn Arg Ser Arg Ala Cys
180 185 190

His Pro Cys Ser Pro Met Cys Lys Gly Ser Arg Cys Trp Gly Glu Ser
195 200 205

Ser Glu Asp Cys Gln Ser Leu Thr Arg Thr Val Cys Ala Gly Gly Cys
210 215 220

Ala Arg Cys Lys Gly Pro Leu Pro Thr Asp Cys Cys His Glu Gln Cys
225 230 235 240

```
Ala Ala Gly Cys Thr Gly Pro Lys His Ser Asp Cys Leu Ala Cys Leu
                245                 250                 255

His Phe Asn His Ser Gly Ile Cys Glu Leu His Cys Pro Ala Leu Val
            260                 265                 270

Thr Tyr Asn Thr Asp Thr Phe Glu Ser Met Pro Asn Pro Glu Gly Arg
        275                 280                 285

Tyr Thr Phe Gly Ala Ser Cys Val Thr Ala Cys Pro Tyr Asn Tyr Leu
    290                 295                 300

Ser Thr Asp Val Gly Ser Cys Thr Leu Val Cys Pro Leu His Asn Gln
305                 310                 315                 320

Glu Val Thr Ala Glu Asp Gly Thr Gln Arg Cys Glu Lys Cys Ser Lys
                325                 330                 335

Pro Cys Ala Arg Val Cys Tyr Gly Leu Gly Met Glu His Leu Arg Glu
            340                 345                 350

Val Arg Ala Val Thr Ser Ala Asn Ile Gln Glu Phe Ala Gly Cys Lys
        355                 360                 365

Lys Ile Phe Gly Ser Leu Ala Phe Leu Pro Glu Ser Phe Asp Gly Asp
    370                 375                 380

Pro Ala Ser Asn Thr Ala Pro Leu Gln Pro Glu Gln Leu Gln Val Phe
385                 390                 395                 400

Glu Thr Leu Glu Glu Ile Thr Gly Tyr Leu Tyr Ile Ser Ala Trp Pro
                405                 410                 415

Asp Ser Leu Pro Asp Leu Ser Val Phe Gln Asn Leu Gln Val Ile Arg
            420                 425                 430

Gly Arg Ile Leu His Asn Gly Ala Tyr Ser Leu Thr Leu Gln Gly Leu
        435                 440                 445

Gly Ile Ser Trp Leu Gly Leu Arg Ser Leu Arg Glu Leu Gly Ser Gly
    450                 455                 460

Leu Ala Leu Ile His His Asn Thr His Leu Cys Phe Val His Thr Val
465                 470                 475                 480

Pro Trp Asp Gln Leu Phe Arg Asn Pro His Gln Ala Leu Leu His Thr
                485                 490                 495

Ala Asn Arg Pro Glu Asp Glu Cys Val Gly Glu Gly Leu Ala Cys His
            500                 505                 510

Gln Leu Cys Ala Arg Gly His Cys Trp Gly Pro Gly Pro Thr Gln Cys
        515                 520                 525

Val Asn Cys Ser Gln Phe Leu Arg Gly Gln Glu Cys Val Glu Glu Cys
    530                 535                 540

Arg Val Leu Gln Gly Leu Pro Arg Glu Tyr Val Asn Ala Arg His Cys
545                 550                 555                 560

Leu Pro Cys His Pro Glu Cys Gln Pro Gln Asn Gly Ser Val Thr Cys
```

565 570 575

Phe Gly Pro Glu Ala Asp Gln Cys Val Ala Cys Ala His Tyr Lys Asp
580 585 590

Pro Pro Phe Cys Val Ala Arg Cys Pro Ser Gly Val Lys Pro Asp Leu
595 600 605

Ser Tyr Met Pro Ile Trp Lys Phe Pro Asp Glu Glu Gly Ala Cys Gln
610 615 620

Pro Cys Pro Ile Asn Cys Thr His Ser Cys Val Asp Leu Asp Asp Lys
625 630 635 640

Gly Cys Pro Ala Glu Gln Arg Ala Ser Pro Leu Thr Ser Ile Val Ser
645 650 655

Ala Val Val Gly Ile Leu Leu Val Val Val Leu Gly Val Val Phe Gly
660 665 670

Ile Leu Ile Lys Arg Arg Gln Gln Lys Ile Arg Lys Tyr Thr Met Arg
675 680 685

Arg Leu
690

<210> 12
<211> 15
<212> PRT
<213> Human immunodeficiency virus type 1

<400> 12
Gly Gly Gly Gly Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg
1 5 10 15

<210> 13
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: internalizing domain derived from HIV tat protein

<400> 13
Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg
1 5 10

<210> 14
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 14
cggtcgacga gctcgagggt c 21

<210> 15
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 15
cagtctccgc atcgtgtact tccg 24

<210> 16
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 16
gagccgcagt gagcaccatg gag 23

<210> 17
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 17
gctgccgtcg cttgatgagg atc 23